DOCUMENT RESUME

ED 214 960 TM 820 111

AUTHOR	Quellmalz, Edys; Snidman, Nancy
TITLE	Evaluation Workshop VI: Summative Evaluation. New Edition
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
PUB DATE	[75]
NOTE	179p.; For related documents, see TM 820 106-110.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Administrators; Data Collection; Decision Making; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Evaluative Thinking; Questioning Techniques; *Summative Evaluation; *Training Methods; *Workshops

ABSTRACT

Summative evaluation is the final stage of the evaluation framework developed by the Center for the Study of Evaluation. Evaluation is a decision-oriented activity. The four-stage evaluation framework consists of needs assessment, program planning, formative and summative evaluation. The handbook presents exercises in summative evaluation. The workshop's purpose is to clarify evaluation and techniques by providing opportunities to discuss and apply them to hypothetical situations. Several areas of evaluation are illustrated including a document file which leads to the exercise in formulating evaluation questions. Measurement considerations determine the means used to answer these questions. The measurement devices must be studied to determine the appropriate instrument. A section of the handbook focuses on selection and construction of tests. When measures to be used to answer evaluation questions are determined, an evaluation design is formulated for each measure. When the components of the evaluation system have been identified, the evaluator may proceed to implement the plan. After the program is completed, the data are analyzed. A written report at the conclusion of the evaluation is presented to the decision makers. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214960

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EVALUATION WORKSHOP VI
SUMMATIVE EVALUATION*

Project Director: Edys Quellmalz

Project Coordinator: Nancy Snidman

Developers: Marcella Pitts
Kathi Yohn

Writers: Darrell Couturier
Rachel Grala
Russell Hunter

Editorial Assistant: Shelby Popham

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 111

*A new edition of the sixth of the CSE Evaluation Workshop series incorporating work begun by Dr. Susan Vivell and by Laura Spooner-Smith. The current version included original work by the project staff

EVALUATION WORKSHOP VI
SUMMATIVE EVALUATION

CONTENTS

Introduction to the Workshop

A. Goals and Objectives A - 1
CSE Evaluation Framework A - 2-5
Formative and Summative Evaluations A - 6-9
Exercise #1: Categorizing Evaluation Activities A - 10-13
Feedback N - 1

B. Introduction to the Simulation B - 1-4

C. Determining the Evaluation Context C - 1-6
Exercise #2: Determining the Context of the Evaluation C - 7-14
Feedback N - 2-3

D. Evaluation Questions D - 1-9
Formulating Evaluation Questions D - 10
Exercise #3: Formulating Evaluation Questions D - 11-13
Feedback N - 4-5

E. Documents File E - 1-12

F. Measurement Considerations F - 1-7

G. Measurement Devices G - 1-6

H. Choosing Tests to Measure Achievement Outcomes H - 1-4
Sample Test Specifications H - 5-8
Exercise #4: Constructing Test Specifications H - 9-11
Feedback N - 6-7

I. Measurement Concerns for Evaluation Questions I - 1-9
Exercise #5: Choosing Measurement Devices I - 10-11
Feedback N - 8

J. Evaluation Design J - 1-11
Exercise #6: Determining Appropriate Evaluation Designs J - 12-16
Feedback N - 9-11

K. Data Collection

Guidelines K - 1-3
Common Problems K - 4-8

L. Data Analysis L - 1-8
Exercise #7: Determining Appropriate Analyses for Evaluation Questions L - 13-20
Feedback N - 12

M. Reporting Information M - 1-19
Exercise #8: Drawing Conclusions and Making Recommendations M - 20-22
Feedback N - 13-14

N. Feedback N - 1-14

Exercise #1: Categorizing Evaluation Activities N - 1
Exercise #2: Determining the Context of the Evaluation N - 2-3
Exercise #3: Formulating Evaluation Questions N - 4-5
Exercise #4: Constructing Test Specifications N - 6-7
Exercise #5: Choosing Measurement Devices N - 8
Exercise #6: Determining Appropriate Evaluation Design N - 9-11
Exercise #7: Determining Appropriate Analyses for Evaluation Questions N - 12
Exercise #8: Drawing Conclusions and Making Recommendations N - 13-14

Selected References

SECTION A

INTRODUCTION TO THE WORKSHOP

SECTION A

INTRODUCTION

EVALUATION WORKSHOP VI

A. Introduction to the Workshop
  1. Goals
  2. Objectives
  3. CSE Evaluation Framework
  4. Exercise 1

Evaluation Workshop VI: Summative Evaluation is designed to enable you to acquire evaluation skills that are essential in a summative evaluation. The evaluator's main responsibility at any stage of an educational evaluation is to provide information to decision makers. In a summative evaluation, the evaluator must determine the overall effectiveness of a fully developed program. Therefore, the evaluator must have the skills to:

1. Specify evaluation questions
2. Develop data collection instruments and design
3. Analyze and report evaluation findings

Objectives

Evaluation Workshop VI: Summative Evaluation is not designed to train people in all of the skills needed by an evaluator in a Summative evaluation. Instead, the workshop will give you a working familiarity with the essential skills for a Summative evaluator. Specifically, the objectives of Evaluation Workshop VI are to train participants to:

1. Identify the procedures of Summative evaluation
2. Identify decision-oriented evaluation questions appropriate for Summative evaluation
3. Select data collection instruments appropriate for evaluation questions
4. Select an appropriate evaluation design for the evaluation questions and program context
5. Identify appropriate methods of data analysis and reporting
6. Identify complete evaluation reports

CSE EVALUATION FRAMEWORK

Summative Evaluation, the topic of this workshop, represents the final stage in a four-stage evaluation framework developed by CSE. Supporting the evaluation framework is one key notion: the purpose of evaluation is to aid the process of educational policy and decision making. With this purpose in mind, evaluators have the twin responsibilities of securing trustworthy information about an educational enterprise and conveying the information -- in a timely and useful form -- to people whose actions will be guided by the evaluation effort.

Evaluation, as it will be treated in this workshop, is thus a decision-oriented activity. Such an orientation leads to the following definition of evaluation, a definition which applies to each of four stages of the Center's framework:

> Evaluation is the process of determining the kinds of decisions that need to be made about an educational enterprise; selecting, collecting, analyzing and reporting evaluation data to aid informed decision-making.

One of the first tasks facing an evaluator is, then, to identify the kinds of decisions that the evaluation is to serve. Only after this is accomplished can evaluators choose and implement an appropriate set of evaluation procedures. The Center's evaluation framework is intended to help evaluators pinpoint the classes of decisions that are to be made; it also should help to guide the selection of an evaluative approach that will yield the needed information. According to the CSE framework there are four major classes of decisions associated with each stage of evaluation.

Each stage, its decision focus and its associated evaluation activities are:

CSE EVALUATION FRAMEWORK

Decision focus	Appropriate Evaluation Activity	Activity Description
Identify "high priority" goals that will serve as the basis for new or revised educational programs.	NEEDS ASSESSMENT	Select from among a full range of alternatives the goals to which resources should be committed. Determine extent to which current program is achieving the "high priority" goals.
Decide the means for achieving previously selected goals.	PROGRAM PLANNING	Select and/or develop the program most likely to attain goals. Develop a plan, including evaluation procedures, for the proposed program.
Decide where improvements in an on-going educational program are needed.	FORMATIVE EVALUATION	Measure program processes and components to ensure program is following the plan. Assess achievement of program goals. Recommend improvements in program where needed.
Determine the overall effectiveness of a fully-developed program to aid decisions about continuing, expanding, or dropping program.	SUMMATIVE EVALUATION	Measure program outcomes and document program processes to permit informed decisions about overall effectiveness.

As this chart suggests there is an orderly and logical sequence to the evaluation stages, beginning with Needs Assessment, proceeding through Program Planning and Formative Evaluation, and concluding with Summative Evaluation. This logic does, in theory, hold true. In actual practice, though, an evaluator rarely has an opportunity to participate in all four evaluative stages for a single program. It is far more likely that services of an evaluator (or several evaluators) will be called for at various points during the life of a program. When an evaluation is called for, it also is rare for the evaluator's mission to be explicityly defined, such as: "We have now completed a formative evaluation; will you please conduct a summative evaluation of our program?"

The ambiguity inherent in many evaluation projects is a powerful reason for evaluators to have a good conceptual understanding of the evaluative activities associated with different phases of educational program development. The following time line relates typical concerns and activities of program development to the services that evaluators might be called upon to provide. Note that the diagram does not indicate an exact proportion of time that should be devoted to each phase of activity. Clearly, for example, the time frame for evaluating a two week instructional program will be considerably condensed compared to a semester-long program.

LIFE OF A PROGRAM AND RELATED EVALUATION ACTIVITIES

	NEEDS ASSESSMENT	PROGRAM PLANNING	FORMATIVE EVALUATION	SUMMATIVE EVALUATION
PHASE OF PROGRAM DEVELOPMENT	·goals/needs that should be addressed are to be identified ·a specific program has not yet been chosen	·high priority goals have been identified ·a method for achieving goals must be selected ·the evaluation of the selected method must be planned	·a program has been selected and put into operation ·program is still "fluid;" modifications will be made	·program has been running for awhile; procedures and methods are fairly stable (no major changes will be made)
WHAT THE EVALUATOR CAN DO	·help to clarify goals ·assist with identifying needs ·measure effectiveness of current programs vis-a-vis important goals	·facilitate planning decisions ·help program planners develop, adapt or select program to meet high priority goals ·build the evaluation component into the program plan	·determine extent to which program is running as planned ·measure program's progress in attaining its goals ·provide recommendations for improvement	·document program processes ·measure attainment of specified goals ·determine whether program is promoting any unplanned effects ·compare to competing program(s)
EVALUATION FINDINGS NORMALLY REPORTED TO	·people responsible for allocating instructional resources (e.g., funding agency; district, state, local administrators) ·people responsible for instructional planning	·people responsible for instructional planning (e.g., teachers, curriculum specialists, administrations) ·community	·program director ·program staff	·people responsible for allocating instructional resources ·groups concerned with program effectiveness
EVALUATION FINDINGS TYPICALLY RELEASED	·when the evaluation is completed	·evaluation plan within program plan submitted to funding agency	·throughout course of the evaluation	·when the evaluation is complete

As this time line shows, it is relatively easy to distinguish needs assessment and program planning from the subject of this workshop -- summative evaluation. Needs assessment evaluation involves determining what goals or objectives an educational program should try to achieve and how well these goals and objectives are currently being achieved. A needs assessment is concluded when a final set of goals is agreed upon. Program planning involves both developing, selecting, or adapting programs to meet the needs that were identified during needs assessment and planning the subsequent evaluation of that program. This evaluation may be either formative or summative.

The distinction between formative and summative evaluation is less clear than is the distinction between, let's say, needs assessment and summative evaluation.

As noted in the chart on page A-3 the decision focus of formative evaluation is to improve a newly developed or newly implemented program. The major activity during formative evaluation is to monitor the program once it is underway to make sure that all its facets have been implemented and that program objectives are being met. The data collection strategy is intensive. A formative evaluator frequently checks on the implementation of the program, whether it is running according to plan and on schedule, whether the right students are getting the instruction as prescribed and so on. Frequent checks of student progress and other program objectives also are made. All of this information is periodically reported by the formative evaluator to the program staff -- teachers and administrators -- so they can correct any problems and know exactly what is happening.

In contrast, the decision a summative evaluation supports is the continuation, expansion or termination of a program. Since the aim of a summative evaluation is to judge the overall effectiveness of an established program, the evaluator may adopt an objective, third part role to avoid charges of bias. Data is collected less frequently than during formative evaluation since the evaluation is designed to document major program processes and performance on major program objectives.

While these distinctions between formative and summative evaluations hold up in theory, in practice the distinctions between them frequently becomes hazy. Actual evaluations vary along a continuum ranging from formative to summative evaluations. Since we frequently hear people referring to a formative or summative evaluation of a program, distinctions between the two stages will be mentioned throughout the workshop.

EXERCISE 1

CATEGORIZING EVALUATION ACTIVITIES

Directions:

The Center for the Study of Evaluation has developed a four-stage evaluation framework. The four stages are:

A. Needs Assessment

B. Program Planning

C. Formative Evaluation

D. Summative Evaluation

During each of these stages, the people responsible for conducting an evaluation focus their attention on certain activities. In this exercise you will be given a list of six situations calling for an evaluation. Your task is to select the evaluation stage which is most appropriate for the situation.

EXERCISE 1

Directions: In the space next to each statement below, please write the letters of the type of evaluation activity which would be most appropriate to the situation described. You may wish to refer to the chart on page A-3 which summarizes the evaluation activities.

NEEDS ASSESSMENT = NA
PROGRAM PLANNING = PP
FORMATIVE EVALUATION = F
SUMMATIVE EVALUATION = S

1. In the last three years, school bond propositions for the Eaglecliff District have failed in the polls. The School Board consequently is faced with making budgetary cut-backs in the district's program. They decided to make the cut-backs on the basis of an intensive evaluation of three well-established instructional programs.

   Evaluation Activity NA PP F S _____

2. The Westchester School District will be granted a considerable amount of funds from the Emergency Assistance Act, if the district can supply clear justification for the manner in which the funds will be spent. Legislation supporting the Emergency School Assistance Act states that "funds are to be allocated for program only in cases where a consensus concerning high priority instructional objectives has been reached among parents and school staff."

   Evaluation Activity NA PP F S _____

3. Normal High School has a large proportion of students who tend to drop out of school at the age of sixteen. Several semesters ago one of the teachers instituted a program in her homeroom with the aim of motivating students to remain at school until receiving their diplomas. The program, which at first operated in an informal way and with few materials, has gradually developed into a systematic, but time consuming, course of study. The school principal wants to know whether the program is worth the investment of time; if so, the principal would like to have the program administered in all homerooms. An evaluator from the school district has been called in to provide the principal with information about the program's effects.

Evaluation Activity NA PP F S ____

4. Last year, the state of Euphoria allocated several million dollars to school districts on the basis of district plans for administering new pre-school and day-care programs. Most of the plans funded by the State were based on very recent research in child development and were therefore experimental in nature. A large scale evaluation effort was consequently commissioned by the state legislature to monitor implementation of the plans and to periodically measure the children's progress in meeting pre-specified objectives.

Evaluation Activity NA PP F S ____

5. Mount Olive Junior College recently installed a language laboratory with tape recorders, workbooks and an extensive library of references. Though the language instructors are willing to change their usual method of teaching so that they can take advantage of the new equipment, they are not familiar with alternative methods of teaching language with adjunct materials. In addition, several instructors are worried that a change in instructional approach will have a negative effect on students' learning. A consultant has been called in to help them make efficient use of the language laboratory and to suggest ways for measuring results of new instructional approaches.

Evaluation Activity NA PP F S ____

6. Bi-monthly meetings of the staff of a new reading program at Lark Elementary School usually end up in controversy. Some teachers complain that the amount of record keeping is so massive that they can not keep track of student progress; others are dissatisfied with the schedule, while teachers with slow readers claim that the program guidelines are working splendidly. The principal decides to call in an evaluator to help solve some of these problems.

Evaluation Activity NA PP F S ____

* * * * * * * * * * * * * * * * * * * *

SECTION B

INTRODUCTION TO THE SIMULATION

SECTION B

INTRODUCTION TO THE SIMULATION

The purpose of Evaluation Workshop VI is to clarify evaluation concepts and techniques by providing opportunities for you to discuss and apply them. During the workshop you will be working individually and in teams to carry out a summative evaluation in the hypothetical metropolitan school district of Westlake. The exercises in the workshop are intended to approximate, as closely as possible, the steps that the evaluator, Dr. Nakamura, follows to plan and carry out the summative evaluation. Many of the exercise forms (in your Participant Notebook) have been designed to resemble working documents that an evaluator is likely to develop or use during the course of an evaluation. You will follow Dr. Nakamura as she conducts the summative evaluation of the Consumer Education Program, this year at Kennedy High School, located in the Westlake School District. Thus, when the workshop is over, you will have a "case study" of an evaluation that you will have conducted.

Throughout the workshop, in order to further clarify the summative evaluation procedures Dr. Nakamura must follow, we will draw from her experiences with last year's summative evaluation of a counseling program in the same school district. Background information about the school district, Westlake, and the two programs, Counseling and Consumer Education, are summarized below.

Westlake Unified School District is located in a suburb of a large urban area. There are 14 elementary schools, three junior high schools, and two high schools. The average daily attendance is 15,000.

In the workshop simulation, you will play the role of Dr. Shelley Nakamura, an evaluator from a local university who is hired by the district from time to time to conduct summative evaluations in the schools. Other characters in the workshop include:

Mr. Philip McNeil, Assistant Superintendent for Instruction

Ms. Betty Washington, District Evaluator

Ms. Irma Gruen, Principal

Mr. Bill Wolfson, Kennedy High School Consumer Education Teacher

Counseling Program. Last year, Dr. Nakamura assisted in the summative evaluation of a counseling program which was developed to meet the needs of the academically-advanced, college-bound high school students.

Two years prior to the summative evaluation, the district had been awarded money by the Division of Special Education of the State Board of Education, for a research and development project involving students in academically accelerated classes. Irma Gruen, the Kennedy High School Principal decided to focus on these students' counseling needs. Betty Washington, the District Evaluator was called in from the beginning to help out in all facets of the development process.

During the Needs Assessment stage, the students, counselors and teachers all agreed that the central objective should be study habits since there was a large discrepancy between actual performance and desired performance.

During the Program Planning stage, Betty Washington and a team of counselors together planned a special group counseling program to teach study habits and to get students to exhibit them in measurable ways, e.g., handing work in on time.

Throughout the formative evaluation, Betty Washington monitored the program to ensure that the program was implemented as planned, measured the progress of students in the program and recommended program changes on the basis of the data collected.

The summative evaluation of this program will be referred to throughout the workshop to help you, in your role as Dr. Nakamura, evaluate the Consumer Education program.

Consumer Education Summative Evaluation. This year, Dr. Nakamura was called in to do a summative evaluation of the Consumer Education Program at Kennedy High School. The development of this one semester elective course is supported with funds from the district set aside for program development.

The course was developed in response to requests from students, parents and several board members that the high school offer instruction in consumer affairs and accounting for personal use.

The Needs Assessment stage indicated that the course objectives not being met by any other program in the high school included: knowledge of consumer related regulatory agencies, and knowledge of how to resolve a consumer complaint, open and maintain a bank account, complete a tax return, buy on cash or credit, and do comparison shopping.

Bill Wolfson, a long-term substitute for the Westlake schools, not only taught this course but was also the course developer. Throughout the Program Planning stage, Bill's developmental work included both identifying instructional materials and strategies, and the creation of measures to determine if course objectives were being achieved. At the outset, it was agreed that the course would be put to a summative test during its third semester of operation.

During the first two semesters, Mr. Wolfson kept careful record of the materials and teaching strategies used as well as of the student performance on progress measures. In the second semester, Mr. Wolfson altered certain first-semester instructional procedures on the basis of the feedback provided by his progress measures and the formative evaluation conducted by Betty Washington.

The program is now entering its third semester and Dr. Nakamura has again been asked to conduct a summative evaluation for the district.

SECTION C

DETERMINING THE EVALUATION CONTEXT

SECTION C

DETERMINING THE EVALUATION CONTEXT

C. Determing the Evaluation Context
  1. Context Questions
  2. Exercise 2

In nearly all newly initiated social projects, such as an educational evaluation, there is a period of activity and that could be called the "start-up" phase. Start-up activities usually are informal, are rarely written into an official evaluation schedule, and, in fact, generally occur before an evaluator has established a formal evaluation plan. Despite the informality, it is at this point that Dr. Nakamura must take the time to gain a thorough understanding of the nature, scope and context of the up-coming evaluation of the Consumer Education Program.

Investigating the contextual factors operating in a school setting involves learning about resources, constraints, social dynamics, political pressures and environmental conditions that might influence decisions about how the evaluation should be conducted. It also means becoming thoroughly familiar with the program itself.

Talking with program or administrative staff, reviewing records and previous evaluation reports on the program, meeting with influential leaders in the community -- these are among the many sources of information about the educational, social, and political context. For example, Dr. Nakamura might learn from a teacher that the program staff perceives the purpose of the evaluation as formative (i.e., to make changes in the program) while administrators see it as summative in nature. Similarly documents in a district file might indicate that the program's objectives have been radically modified since the original program plan was developed.

It is a good idea to share and verify various contextual factors with the individual(s) who commissioned the evaluation, and to do so before drawing up detailed evaluation plans. In this way both parties -- evaluators and their employers -- will hold from the start a common view of what the evaluation is expected to accomplish and what it realistically can accomplish.

Outlined below are some "start-up" questions that evaluators might try to answer when they first become involved in an evaluation project. Please note that the list is suggestive, but not inclusive, of all concerns that may be important in different settings.

Context Questions

I. Program Characteristics

A. What is being evaluated and why? - overall description of the program.

1. What is the subject area?
2. What is the origin of the program?
   a. How did it get started?
   b. Who were the prime movers?
   c. What was the major impetus for the program?
3. What are the general goals? How were these established? (Needs assessment, mandated, other)
4. What are the specific objectives? How were they selected? (Research, theory, rationale)

B. What strategies (teaching style or instructional methods are included to bring about program goals?

C. What are the program's materials, basic procedures and activities?

D. What kinds of students does the program serve?

   1. for whom was the program planned?
   2. for what age, grade level, ability level is the program appropriate?
   3. how are the students selected for a program?
   4. what criteria, if any, determine their initial and continuing participation?

E. Who are the individuals operating the program?

   1. how many teachers, administrators, secretaries, clerks, aides, volunteers and others are active in running the program?
   2. did they receive special training?
   3. how much time per week do they devote to the program?

F. How long is the program?

G. What is the total cost of the program and its allotted budget?

II. Evaluation Requirements

A. What is the general purpose of the evaluation?

   1. what is the place of evaluation in the organization? Is it to be conducted at the local, district or state level?
   2. what questions should the evaluation answer?
   3. what is the scope of the evaluation?
   4. what decisions about the program will be made based on the evaluation?

B. How much time is allotted to conduct the evaluation?

    1. what are the reporting deadlines?
    2. how much time is available for data collection?

C. Are there mandated evaluation requirements to comply with funding agency guidelines and/or laws?

D. What kind of evidence of program success is either required, desired or possible?

    1. what kinds of potentially useful data are already available? (previous evaluation reports, test scores, teachers' progress reports, current materials, etc.)
    2. what kinds of potentially useful data can be obtained? (interviews, on-site observations, special performance tests, questionnaires, etc.)
    3. Are there any problems with the data on hand? (e.g., test administration in which lower ability students were excused, unwillingness of administrators to release information?

E. Are there any design constraints?

    1. is a certain sample size necessary? Is it available?
    2. is random assignment of students possible?

F. What is the allotted budget for the evaluation?

    1. what monetary and staff resources are available?

G. What audiences are there for information about the program and the evaluation?

    1. what are the audiences for interim program information and the final evaluation report?
    2. how frequently should reports be made?

III. School Organizational Structure

A. Who are the main decision makers connected with the outcome evaluation?

   1. why are they involved in the evaluation? (self-initiated or required)

   2. what are their motivations? (e.g., to determine if a program is successful vs. showing that a program is successful)

   3. what is their knowledge of evaluation?

   4. what is their commitment to/attitude toward evaluation?

   5. do they agree with the proposed requirements of the evaluation?

   6. what is their managerial style (shared, authoritarian)

B. Who does the evaluator work with and report to?

C. What are the formal and informal lines of communication within the program?

   1. who are the opinion leaders in the program?

D. What is the status of the program's administrative group?

   1. how strong is their support in the community?

   2. how strong is their evaluation expertise?

E. Are there other special interest groups in the school structure with involvement in and attitudes toward the program and its evaluation?

IV. Community Organizational Structure

A. What involvement with the program do parents, businessmen and other community groups have?

   1. what information about the program and evaluation does the community have?

   2. what attitudes toward the program and evaluation does the community have?

B. Who are the potential critics of the program and/or the evaluation?

   1. what are their concerns?

C. What "political" conditions in the community, district or state exist that may influence program and evaluation activities (e.g., Is a changing school enrollment creating need for expanding or deleting programs? Is a local newspaper editorializing for local school reforms?)

V. Role of the Evaluator

A. Is the evaluator internal or external? (part of the program staff or a third party outsider)

B. Is the evaluator to be helpful and nurturing, or neutral and objective?

EXERCISE 2

DETERMINING THE CONTEXT OF THE EVALUATION

Shelley Nakamura is keenly aware of the importance of attending to the contextual factors of the Consumer Education Program evaluation. Therefore, she has decided to contact Betty Washington, the Internal District Evaluator, to fill in the background information about the program. A transcript of their conversation is on pp. D-10 - D-14.

In this exercise, you will be using the information in the conversation between the two evaluators. (Answer Dr. Nakamura's questions about the program.) Each team member will be responsible for completing one group of contextual questions.

Team Member A: Concerns about Program Characteristics
Team Member B: Concerns about Evaluation Requirements
Team Member C: Concerns about School Organizational Structure

Exercise 2

SECTION C

Determining the Evaluation Context

For Team Member A: Concerns About Program Characteristics

1. What is the most up-to-date source of information about the program's operation and goals?

2. What activities do the staff feel are contributing most to program success?

3. Who are the individuals operating the program?

For Team Member B: Concerns about Evaluation Requirements

4. What unexpected positive effects does the program seem to promote? What negative effects?

5. What kinds of evaluation information are available or obtainable?

6. What monetary or staff resources are there to conduct the evaluation?

For Team Member C: Concerns about School Organizational Structure

7. What are the decision makers?

8. What are the levels of authority and communication within the program?

9. Who are the potential critics of the program or of the evaluation?
What are their concerns?

Determining the Context of the Evaluation

Characters: Shelley Nakamura (N) and Betty Washington (W)

N: Betty, good morning. Thanks for setting aside some time for me on such short notice.

W: Good morning, Shelley. No problem; the summer is not nearly as hectic as the regular school year.

N: As you know, the assistant superintendent for instruction wants me to conduct an evaluation of the Consumer Education Program at Kennedy High School this fall.

W: Yes, I pulled my file on the Consumer Ed Program which contains the formative evaluation report done last semester, correspondence, and an updated version of the program plan, and had copies made for you.

N: Good, thanks. I will go over these very carefully. Since you were involved with the program last year, maybe you can give me some insights that will help me get started.

W: Sure. What can I tell you?

N: Well, first, tell me who has been involved in running the program.

W: Well, there really aren't that many people involved; the program teacher, Bill Wolfson, and the high school principal, and several parents who have been helping arrange field trip sites. The Assistant Superintendent of Instruction is in charge of all experimental programs, and of course, I work for him.

N: Oh, so everyone reports to the Assistant Superintendent?

W: Yes. There are so few people involved, it's pretty straight-forward.

N: Any staff problems you have observed?

W: No. Bill Wolfson has great enthusiasm for the program, but then he was a long-term substitute hired specifically to develop and teach the course on a trial basis. Bill really wants to see the course continued; this year he has been teaching three consumer education classes, and the rest of the day he was teaching release-time courses for the chairman of the social studies department who took part in a district program. If the course is not continued beyond this year, Bill will most likely have to go back to long-term substitute status. There's one other thing about the staff. I think there might be some annoyance on the part of the other teachers. The field trips have caused absences in their classes, and that makes more work. If you need any of the other teachers for the evaluation, be sure that you do so at their convenience as much as possible.

N: That's a good point, thanks. Aside from teachers being unhappy about class disruptions, are there any real critics of the program, or is there anything unusual going on right now that might influence the program or our upcoming evaluation activities?

W: I wouldn't say that there are any real critics of the program; it's kind of hard to be against teaching high school students to be informed consumers. The only criticism--and it's not just leveled at this particular program--has to do with cost. The last two or three people elected to the school board are very concerned about money. As you know, the last bond issue failed, and that hurts even more. Some board members probably feel more pressure to keep costs down than to respond to students and parents who want this type of program.

N: And the Assistant Superintendent is looking for a fully documented evaluation report no later than the last week in March.

W: That's right. The purpose of the evaluation is to provide information on the success of the program and whether it should be continued at Kennedy High and started at Maples High. The Assistant Superintendent will be making that decision, probably after consultation with the high school principals involved. And they want to make the decision in plenty of time to get it implemented next year, so timely completion of the evaluation report is important.

N: Yes, that being the case, I'm glad we've got a little bit of time before school opens in the fall to get this evaluation planned. Do you think I'll have any problems collecting information this fall? For instance, will I be able to observe the program, give out some questionnaires to students who are in the program or were in the program?

W: You can certainly go in and observe. Bill would be especially pleased to have you watch the Consumer Beware Game he developed, which he feels makes students apply the concepts just taught in the practice exercises. Bill and the program students would also be willing, I'm sure, to fill out questionnaires. For students who are not now in the program, you will have to devise a system for reaching them that does not interfere with classroom activities--their homerooms are probably the least obtrusive way to do that.

N: Oh, good, so I will have some access to students who are not in the program currently?

W: Yes. More students wanted the Consumer Education class this fall than we had room for. The popularity of the course may be caused by a few students who took Consumer Ed. last semester. They began to write a very flamboyant column in the school newspaper about student's legal rights, especially the right of students to control their own records and files. There is a waiting list for the spring semester you might want to do something with. We also have the names of all of the students from last year, although some of them were seniors and graduated. Remember that you must get clearance from the Assistant Superintendent for any measures you devise for students to fill out. And remember to file all of your raw data with him, too.

N: Yes, of course. I would like you, the Assistant Superintendent and Bill not only to approve them, but also to suggest whatever revisions seem necessary.

W: I'll be happy to go over whatever you want me to.

N: Good, thanks. I'm definitely counting on your consultation along the way. The Assistant Superintendent also said that in addition to your helping me as needed, your office would also handle all of my clerical and data processing needs.

W: Yes, that is correct. Is there anything else we can do for you this morning?

N: No, I don't think so. Thanks for getting me off to a good start. I'll be in touch with you again after I've had a chance to go over what you've given me in your Consumer Education file.

SECTION D

EVALUATION QUESTIONS

SECTION D

EVALUATION QUESTIONS

D. Evaluation Questions
  1. Outcome
  2. Implementation
  3. Cost
  4. Research
  5. Unanticipated Outcome
  6. Formulating Evaluation Questions
  7. Exercise 3

In the last section, Shelley Nakamura, as the external evaluator in the simulation, investigated the contextual factors for the Consumer Education Program. Now that she has all the necessary background information on the program, resources, and people, Dr. Nakamura is ready to start planning the evaluation. To conduct an effective summative evaluation and ensure that the data to be collected yields the desired information, Shelley plans to organize the evaluation around three central evaluation questions:

OUTCOME QUESTION: Did the program meet its objectives?

IMPLEMENTATION QUESTION: What does the program look like in operation?

COST QUESTION: What costs, including dollar expenditure, time, resources were necessary for the program?

Depending on the context in which the evaluation is taking place, a summative evaluation might also be concerned with two other categories of questions:

RESEARCH QUESTION: How do specific program processes relate to program outcomes?

UNANTICIPATED OUTCOME QUESTION: What other outcomes can be attributed to the program?

The evaluation questions which are particular to each program should be arrived at jointly by those commissioning the evaluation effort (or their representatives) and the evaluator. In this simulation, Mr. McNeil, Betty Washington and Bill Wolfson together with Dr. Nakamura, should choose and/or write these questions with the awareness that the questions must be both consistent with information needs of the decision makers and compatible with whatever resource and social constraints affects the evaluation.

Evaluation questions are especially important because they give direction to all of the activities that follow. They guide the selection of data collection instruments, evaluation design, and procedures for data analysis. They also impose an efficiency on the subsequent steps in the evaluation process, because they identify the kinds of information that must be collected and the kinds of information that are superfluous to the needs of the decision makers.

OUTCOME QUESTIONS

Did the program meet its objectives?

For many programs, most or all of the program objectives are concerned with people and the way the program affects their behavior, knowledge, feelings, and so on. (Some programs do have objectives that are not related to changes in people, e.g., cost factors; how an evaluator should treat such "non-people" objectives is discussed in a later section of the Workshop.) Typically the objectives of a program will fall into three basic categories: cognitive, psychomotor and affective.

Though the objectives of a program often focus on students, the objectives may be for any other group, such as parents, teachers or aides, who are expected to be affected by the program in some significant way. For instance, a program might be planned to increase teacher interaction with parents or to make aides more proficient in using a particular instructional approach.

The ease with which evaluators arrive at a final listing of questions about objectives will depend largely upon the kind of program planning that has been done. The task of specifying the questions may be relatively simple if the program staff or developers already have operationalized the precise skills, knowledge, attitudes, etc. that participants are expected to acquire.

On the other hand, if the intended outcomes of the program have been formulated only as rather general goals, then the evaluator will need to refine the goal statements into more precise descriptions of desired attainments. These descriptions will, in turn, serve as the basis for evaluation questions. For example, there were a number of different study skills components to the counseling program at Kennedy High School. The outcomes for one of those components are listed below and illustrate how general goal statements might be refined into behavioral objectives.

General Goal 1	Behavioral Objectives
1. Students will schedule time effectively for classes, athletics, study and recreation time.	1.1. Given a blank scheduling sheet, students will fill in times for their classes, study, and recreation.

CSE
ARCHIVE COPY

2. Students will task analyze major class projects and complete them in an organized manner.	2.1. Given a major class project designed to be completed over several weeks, students will break the task down into components and provide due dates for each component.

Especially when goals of a program are formulated in a general way (like those in the left-hand column above), it is critical that agreement is reached regarding the precise meaning of the goal. It is often the case that generally stated goals take on new and different meanings as a program develops and can be measured in a number of ways. For example, the general goal that students will improve their study habits, the goal which was initially listed for the counseling program developed at Kennedy High School, can mean any or all of the following:

1. That students will follow a regular study schedule.
2. That students will begin assignments on time.
3. That students will stick to a task until complete.
4. That students will hand in assignments on time.

Rephrasing general goals in operational terms, as in the right hand column on the previous page, reduces ambiguity and enables the evaluator to phrase evaluation questions precisely. For example, No. 1, immediately above, translated into the Behavioral Objective on the previous page will enable the evaluator to ask the question, "Did students in the program follow a regular study schedule?" and provide behavioral measures of the outcome. Naturally, an evaluator should ask program staff members and others involved in the program to review the evaluation questions to make sure that they are, in fact, an accurate interpretation of the program's objectives.

If there are similar programs in existence, the decision makers may want to know if an alternative program is a better means of achieving the same or similar objectives. When this is the case, a comparison will be made between two or more programs, and the question becomes, "What are the outcomes of Program X compared with the outcomes of Program Y?" Information about program outcomes is much more powerful when provided along with data on the amount of success of other alternatives.

IMPLEMENTATION QUESTIONS

> What are the key characteristics of the program? and What does the program look like in operation?

Program implementation is necessary for program certification and program replication. If program goals are achieved and it can be proven that the outcomes were caused by the program, implementation will suggest which procedures caused the outcomes. A major goal, in fact, of many agencies is to identify successful programs and replicate those programs in other settings -- be definition replication calls for procedures to be known so they can be duplicated.

Implementation questions call for two basic categories of information: background information and information about program processes.

A. *Background information*. What are the key characteristics of the program? This category investigates the context in which the program operates by focusing on the following components:

- *Origin*. How did the program get started? Who were the prime movers? What was the major impetus for the program?
- *Objectives*. What was the program designed to accomplish? On what basis (research, theory, rationale) were the objectives selected?

- Individuals served by the program. For whom was the program intended? For what age, grade level, ability level, interest group, etc. is the program appropriate? What criteria determine their initial and continuing participation?
- Individuals operating the program. How many and what kinds of teachers, administrators, secretaries, aides, clerks, volunteers and others are active in running the program? Did they receive special training? How much time per week do they devote to the program?

B. Program Processes. What does the program look like in operation? Evaluators need to be selective when they ask questions about processes. Although processes are the means by which goals are supposed to be met, it would be senseless for an evaluator to investigate the hundreds of things that people do while a program is going on. Therefore, it is suggested that evaluators focus on those processes that are expected to have the greatest influence on planned outcomes, are unique to the program, or are most costly (in terms of time and money).

The term "processes" covers a wide range of things. It refers, for instance, to the activities of students, teachers, and others involved in the program; instructional strategies employed; materials and resources used by program participants; basic program procedures. To decide which processes ought to be included in program documentation, evaluators might ask, "What would an observer see students, teachers, or others doing during a typical session of the program?" "What special activities, like a single field trip or lecture, are viewed as critical attributes of the program?" For example, in the evaluation of the Counseling Program, Dr. Nakamura wanted to look at the program process of training students to schedule their time. She looked at their original and revised schedule sheets, along with the class record sheets used by the counselors to provide students with evidence about the effectiveness of their time management.

COST QUESTIONS

> What costs, including dollar expenditure, time and resources were necessary for the program?

It is possible for a program to have objectives that are not related to changes in people but focus instead on the manner in which the program is managed. It is quite likely that these are not publicly announced goals but are instead standards that administrators use to keep the entire school and individual programs running smoothly. Administrative goals or questions generally relate program process to program costs. For example, administrative evaluation questions might be

- Does the time counselors spend in assisting with the new study habits program keep them from performing their regular school assignments?
- What is the opportunity cost of implementing the counseling program for study habits at both district high schools?

Sometimes decision makers will want the evaluation to answer the question: What does the program cost? "Cost" is a word that takes on different shades of meaning depending upon the persons using it. Cost can refer to dollar amounts spent on materials, salaries, construction and other budget items; it can also refer to the allocation of personnel and the wear and tear on program staff.

Cost-type questions about program elements should not be asked in the absence of research questions that relate program components to outcomes. Decision-makers are likely to treat a costly program component differently if they know that it is crucial to the achievement of program objectives. Cost questions are, by definition, tied to decisions about the value of a component, or even an entire program. These decisions can become less subjective when evaluators systematically research the relationship between specific program characteristics and outcomes and report the results.

RESEARCH QUESTIONS

How do specific program processes relate to program outcomes?

Research questions involve a planned variation in the program processes to determine what differential impact, if any, the variation has on attainment of program outcomes. For example, a decision-maker may want to know whether small group discussions or individual workbooks are more efficient at producing certain program objectives. While such information may be useful for program improvement, the focus of the evaluation must be to collect information what will help answer the basic question -- "How useful is this program?"

The research questions that might be embedded in a summative evaluation are not formative evaluation questions used over again. In a formative evaluation, many program conditions may be changed simultaneously as the program is initially implemented and pilot-tested. When this is the case, it is difficult to isolate the impact of any single program component. Keep in mind that the critical characteristic of a research question is that all other program conditions must remain the same while some program components are *systematically* varied.

UNANTICIPATED OUTCOME

What other unplanned outcomes can be attributed to the program?

This question calls for an identification of outcomes -- either positive or negative -- that can be attributed to the program but which were not a part of the original program objectives. A consideration of *all* program outcomes gives decision-makers a more thorough picture of just what the program does accomplish than would an evaluation effort that measures only planned outcomes.

During the operation of any educational program, there is always the possibility that outcomes may occur which were not anticipated, but that nevertheless have great impact. For example, students may master cognitive objectives, but may develop an aversion toward the subject matter. Unanticipated outcomes may be positive or negative. They may be limited to the program being evaluated; they may affect other programs; or, they may affect the school program in general.

Sleuthing for unanticipated outcomes or additional information may pay off in significant findings, but there is also the risk of time being spent on the collection of data that is either unrelated to the program or related but fairly trivial. Just how much digging around is done in search of the unexpected will be determined by the purpose and scope of the evaluation. For example, a time-constrained evaluation for the purpose of making a year-end report to the school district allows much less flexibility than a longer evaluation in which one of the concerns of those commissioning the evaluation is to discover all of the outcomes and spin-offs that can be reasonably attributed to the program in question.

The process of determining whether the planned program objectives have been achieved is a sophisticated process involving selection or development of valid and reliable measures and a research design that controls for as many extraneous variables as possible. Any leads which seemingly point to unanticipated program effects must be as thoroughly and systematically pursued as the planned outcomes to avoid the pitfall of making claims about the program that are based on partial evidence.

FORMULATING EVALUATION QUESTIONS

The language used to phrase evaluation questions must be understood by the decision-makers and others involved in the evaluation. Ideally, everyone involved will be able to look at a list of evaluation questions and agree on what the question means. Clearly stated questions will help the evaluator determine what types of information should be collected, when, and from what respondent group(s).

An evaluation question can be stated with a criterion (e.g., "Did the program cost per student stay within $10.00 of the cost of the regular program?") or can simply be a call for general information (e.g., "How much did the program cost per student?"). Whether or not criteria are included in the evaluation questions will largely be a matter of whether or not the goals of the program are stated in operational terms.

The precise number of evaluation questions is limited by the amount of relevant information decision-makers need and the constraints within which the evaluation must operate (e.g., time, money, resources, and personnel). There may have to be some compromises between the evaluator and groups concerned with the program and the evaluation in order to keep the length of the list manageable.

In the next exercise, you will be asked to play the role of Dr. Nakamura and formulate a set of evaluation questions.

EXERCISE 3

FORMULATING EVALUATION QUESTIONS

In this exercise you will formulate some of the evaluation questions Shelley Nakamura might ask in her evaluation of the Consumer Education program. A series of documents follow the exercise. The information contained in the documents will help you formulate evaluation questions in the exercise. These documents are excerpts from the evaluation folder on the Consumer Education Program which Betty Washington has sent to Dr. Nakamura. They were compiled during the program's pilot year of operation. Information that Dr. Nakamura gathers by reading these documents and by meeting Bill Wolfson, Betty Washington and others will help her formulate a series of evaluation questions to guide the evaluation of the Consumer Education Program.

Directions:

On the next pages is a letter from Shelley Nakamura to Philip McNeil written immediately after Dr. Nakamura read the contents of the Consumer Education folder and had talked with Kennedy High School principal, Irma Gruen and the Consumer Education Program teacher, Bill Wolfson. Read the letter and the exercise to familiarize yourself with the objectives of the exercise. In order to complete the exercise, read the documentation material for the Consumer Education Program that follows the exercise. When you have completed reading these materials, try to formulate at least one question for each of the five categories of evaluation questions in the exercise. (One possible question is already provided.)

EXERCISE 3

August 13, 19XX

Mr. Philip McNeil
Assistant Superintendent of Instruction
Westlake Unified School District

Dear Phil,

In preparation for the evaluation of Kennedy High's Consumer Education program, I have spoken with Betty Washington, Bill Wolfson, and Irma Gruen as you suggested. I have also familiarized myself with Betty's files on the program and am looking forward to all of us meeting on the 25th to draw up detailed plans for the evaluation. In advance of that meeting, let me share with you my understanding of the task at hand.

I think we are all agreed that the purpose of the evaluation is to provide you with data about program outcomes and program processes so that you can make a decision on whether to discontinue, continue, or expand the program. My final report is to be submitted to you no later than March of next year.

As you know from last year's evaluation of the counseling for study habits program, my information gathering will be guided by a set of evaluation questions. In our upcoming meeting, I hope that we will be able to arrive at a set of questions that will give you the kind of evidence you need to make informed decisions about the program. Judging from my various conversations and from the existing program documents, it appears that you may want answers to some or all of the following questions:

1. PROGRAM OUTCOMES "Is the program meeting its objectives?"

   1) Are students able to list the steps to follow in resolving consumer complaints?

   ____

   ____

2. PROGRAM IMPLEMENTATION "What are the critical program characteristics?"

   1) Is the Consumer Education game being used as planned?

   ____

   ____

3. COST QUESTIONS "What costs, including dollar expenditure, time, resources, were necessary for the program?"

1) What is the cost of running the program compared to other one-semester elective social science courses?

4. RESEARCH QUESTIONS "How do specific program processes relate to program outcomes?"

1) What is the effect of field trips on achievement of the course objectives?

5. UNANTICIPATED OUTCOMES "What unplanned outcomes can be attributed to the program?"

1) Is the recent campus interest in student rights related to the Consumer Education class?

The above list is certainly tentative in my mind, but hopefully will serve as a basis for discussion. I look forward to your input on a final set of questions on the 25th so that we can get the evaluation underway.

Sincerely,

Shelley Nakamura, Ph.D.

cc: Irma Gruen
Betty Washington
Bill Wolfson

SECTION E

DOCUMENT FILE

SECTION E

DOCUMENT FILE

On the following pages are excerpts from Betty Washington's evaluation folder on the Consumer Education Program in its pilot year of operation. After your team has considered these documents, complete Exercise 3 "Formulating Evaluation Questions" (pp. E-12-14).

EVALUATION FILE ON CONSUMER EDUCATION PROGRAM
WESTLAKE UNIFIED SCHOOL DISTRICT

Consumer Education Program

Revised Plan - June 19XX

Origin. The course was developed in response to requests from students, parents, and several board members that the high schools offer instruction in consumer affairs and accounting for personal use. As well as teaching the course, Bill Wolfson was the course developer. His developmental work over the course of one summer was supported with funds from the Office of the Assistant Superintendent for Instruction. It was agreed between the program funders and the program developer that a formal outcome evaluation would take place during the program's third semester of operation.

Goals. The course objectives are stated in two broad categories, Consumer Awareness and Personal Accounting:

Consumer Awareness Goals

- the student will be able to list steps to follow in resolving a consumer complaint
- the student will list major local, state and federal consumer-related regulatory agencies and identify the functions of each

Personal Accounting Goals

- the student will demonstrate how to open and maintain a personal checking and savings account
- the student will complete a state and federal income tax return
- the student will list the advantages and disadvantages of buying on credit

- the student will demonstrate a knowledge of comparison shopping techniques using price, quantity, and quality as variables
- students will demonstrate a willingness to be planful in handling their personal finances

Characteristics.

The course is one-semester (20 weeks) and meets five times per week during regular school periods of 50 minutes each.

Special materials include:

- "Consumer Beware!" game adapted from a popular parlor game. The game is played several times a week and requires students to apply the skills they have been learning.
- Six consumer education filmstrips produced by the Food and Drug Administration, Federal Trade Commission, and Security and Exchange Commission. The filmstrips cover such topics as determining the quality of a product; calculating a standard unit cost for different package quantities, consumer protection laws and agencies, the function of banks, and different types of credit.
- Two speakers from the community. One speaker belongs to a consumer advocate group and explains what issues the group is currently persuing and what courses of action they are taking. The other speaker is a manager from the local credit bureau who talks about how individuals are assigned credit

Bill is really excited about the game. He developed it himself and feels that it is the key to students applying the course concepts correctly. He [illegible] [illegible] to the [illegible]
JM 6/23

ratings and how credit can work for or against people.

- Three half-day field trips to a bank, car dealership, and supermarket.

Student evaluation materials include:

- knowledge of concepts
- performance tests of skills (e.g., balancing a checkbook, completing a tax return)
- a term paper in which students select a current issue being handled by a regulatory agency, trace the history of the issue; describe the current status of the issue; and predict the resolution or nonresolution of the issue.

Instructional Materials:

- A high school supplementary social studies textbook, American Consumer
- A variety of informational brochures from government agencies and private enterprise.

Instructional Strategies:

- Lecture, class discussion, small group practice (Consumer Beware! game).

Students: The course was designed for high school students (10th grade to 12th grade) of average ability. The course is an elective with priority for enrollment given to upper classmen. Student may not repeat the course unless they received a failing grade.

Faculty: The course is taught by Mr. Bill Wolfson (who also developed the course). Mr. Wolfson has a B.A. in economics and an M.A.T. from State University.

July 17, 19XX

TO: Betty Washington

FROM: Philip McNeil, Assistant Superintendent for Instruction

RE: Evaluation of Consumer Education Program

As agreed last summer when we first made funds available for development of the Consumer Education course at Kennedy High, an evaluation report on the program is due in my office no later than March of this school year. As you know, we want to make a timely decision about continuing the program at Kennedy and, if continued, whether to offer the course at Maples High as well. Shelley Nakamura will be back to conduct the evaluation, and I want you to give her as much assistance as you can. When you discuss the evaluation with Dr. Nakamura, please share with her my concern about knowing if the objectives of the course are being met and if costs associated with the course are in line with what we spend on other social studies electives.

Course Objectives--Average Student Performance

Objective	Percentage of Students Mastering Objective Fall	Percentage of Students Mastering Objective Spring
Consumer Awareness:		
1) steps to follow in resolving a consumer complaint	63	68
2) knowledge of major local, state, and federal consumer-related regulatory agencies	57	65
Personal Accounting:		
3) opening and maintaining checking and saving accounts	87	94
4) completing a state and federal income tax return	83	85
5) knowledge of advantages and disadvantages of buying on credit	81	82
6) knowledge of comparison shopping	92	95

TO: Celia Moore, Social Studies Chairperson

FROM: Jack Whittier, Registrar

RE: Enrollment for Consumer Education Class

Thought you might be interested to know that in addition to filling both sections of the Consumer Ed class for the spring semester, we easily had enough requests to have filled six classes. This thing appears to have really caught on after only one semester. Do you want to schedule more classes for next fall?

Jack –
I have talked with
Bill Wolfson, and we
have agreed to add
a third class period
of Consumer Ed for
next fall
thanks
– Celia

FROM THE DESK OF: Meredith Hunt
Student Publications Supervisor

Bill —

Have you seen our new "Students' Rights" Column in the school paper? Bill Winslow and Terri Gompers who have taken it on claim they got the idea from your class on Consumerism. They said they figured students were also consumers (of education) with rights. Isn't it nice when kids take something they've learned and turn it into something else that's useful?

Cheers —

Meredith

August 23, 19XX

Mr. Bill Wolfson
Kennedy High School
1432 Cloverdale
Conejo, California

Dear Mr. Wolfson,

Following up on my promise to let you know how useful the Consumer Ed course has been in my switch from high school to living away from home and supporting myself, I have two general comments:

1) all of the personal accounting things have really been useful; I'm routinely doing things like saving check stubs and have opened several charge accounts to establish credit. All the stuff on consumer affairs has not really been useful.

2) there are so many small expenses involved when you are living away from home that I think it would be helpful to teach students how to budget their income so that the fixed expenses get taken care of--I really blew my first two paychecks. It would also benefit a lot of people to know about automobile insurance, the same as we learned about health insurance and home insurance.

Hope this helps. Thanks for all your help.

Sincerely,

Connie Norwick

Bill- Help!

Your first two field trips for Consumer Ed have been a disaster as far as other teachers are concerned-- and I understand that there are going to be two more before the semester is over. Don't feel bad--even the smoothest field trip causes disruption when 5-10 students are being pulled from other classes. But can you keep the number of trips down?

Lillian Walker

August 18, 19XX

TO: All Building Principals

FROM: Accounting Office, Triunfo Unified School District

RE: Cost Increase for Field Trips

In order to avoid increased costs for our regular bus transportation to and from school, the price of busses for field trips will be increased by approximately 30% effective September 1.

SECTION F

MEASUREMENT CONSIDERATIONS

SECTION F

MEASUREMENT CONSIDERATIONS

Now that a set of evaluation questions has been identified (i.e., outcome, implementation, cost, research and unanticipated outcome questions), Shelley Nakamura's next task is to select the means by which these questions will be answered. The role of measurement is indeed critical, since the measurement instruments underlie the technical credibility of evaluation results. The relationship between poor instruments and the results they yield can be summarized by a phrase borrowed from computer programming, "garbage in--garbage out!"

There is a rich and sophisticated technology associated with the construction of good test and measures, but an evaluator need not be an expert psychometrician to select credible methods to gather data and avoid the "garbage syndrome." There are several guidelines, described below, that evaluators should follow with all data collection instruments:

- Make sure the measurement device is directly linked to the evaluation question for which it is being used.
- Give consideration to alternative measurement strategies.
- Make sure each measure is appropriate for its recipients.
- Measurement instruments should be as valid and reliable as possible.
- Make the data collection as unobtrusive as possible.
- Instruments should be sensitive to positive as well as negative outcomes.

Make sure the measurement device is directly linked to the evaluation question for which it is being used.

There is always the temptation to collect information merely because it might prove "interesting," or to use a particular measurement device because it is routinely administered at a school site anyway. Unless results obtained from the former or the latter promise to be of significant help in answering evaluation questions, they should not be used for evaluation purposes.

In short, evaluators should carefully examine all potential evaluation measures-especially published instruments--to be certain that there is a match between the measure and the specific information needs. For example, a measure labelled "Grade Two Basic Math Computation Test" might contain many work problems, whereas the evaluation question called only for assessment of simple numerical calculation skills (10x10=_____). Mismatches like this one between what an instrument measures and the intent of evaluation questions, can be unfair to the program and can also damage the credibility of the evaluation. Even if an instrument is entirely suitable in content, it is nonetheless possible for the results to be ambiguous or misleading. This problem is becoming increasingly apparent in achievement testing and has been captured in the debate over the relative merits of criterion-referenced tests versus norm-referenced tests.

The distinguishing feature of good criterion-referenced tests (CRTs) is that they are designed to measure clearly delineated skills, knowledge, or abilities. Results from CRTs indicate an examinee's level of achievement for each area assessed in the test. In other words, the results show an individual's level of mastery for specific skills or abilities.

In contrast, norm-referenced tests (NRTs) compare an examinee's performance with that of a large group of other people, the "norm" group. Results from such measures indicate a person's relative standing in comparison to the norm group. Scores often are expressed as percentiles, grade equivalent scores or standard scores. NRT results provide relative scores describing how a person performed in a general skill area, such as reading comprehension.

When mastery of specific skills or knowledge is to be inferred from test performance, a criterion-referenced test is preferable. They can be obtained from measurement specialists who develop CRTs (also called "domain-referenced tests" or "objectives-based tests"), or they can be developed locally by school personnel.

If data comparing individual students to an "average" student's scores are needed, then a test which provides relative scores should be used. If a norm-referenced test is to be administered, be certain that the group on which the test was normed is comparable to the group which is to be tested; if not, the results may be unfair. Information on the norm group is usually published in the test manual.

Give consideration to alternative measurement strategies.

Because evaluations generally are concerned with assessing a full range of program characteristics including activities and materials, time and costs, attitudes, and, of course, achievement, a range of measurement approaches may be needed. Pencil-and-paper measurement instruments, such as the traditional achievement-type test or quiz, are by no means the sole method for collecting evaluation data. Options open to evaluators, depending upon the nature of the evaluation question, include such strategies as observations, interviews, questionnaires, performance tests, records, checklists and so on; such strategies can be used to obtain information about and from students, teachers, parents and other sources. Evaluation findings are strongest when based on corroborating evidence from several different instruments, especially when different types of instruments (observation and interview) are used with different respondents (students and teachers).

Make sure each measure is appropriate for its recipients.

Whatever measurement devices are selected, each device should be appropriate for those taking it. For example, an interview would obviously be maore appropriate for non-readers than a questionnaire. Again, the credibility of results can be challenged if it appears that the method of measurement was inapprpriate when program goals did not put objectives into any time frame).

Measurement instruments should be valid and reliable as possible.

The terms "validity" and "reliability" are used frequently by test and measurement specialists, and for good reason: validity and reliability are critical attributes for all measurement instruments, whether the instrument is a clock, a high powered electronic sensor, a personality inventory, or a reading test. The concepts of reliability and validty can take on relatively complex technical and statistical meanings when they are applied to the process of test construction, a topic beyond the scope of this workshop. However, they also have faily common sense interpretations with which evaluators should be familiar when choosing measures.

Reliability refers to the consistency with which a measure assesses whatever it is intended to measure. One way to measure can be reliable is over time. For example, a group of students might take an algebra test on Monday. Provided that they receive no practice or instruction on the class of problems included in the test, the group should receive approximately the same score if the same test were to be administered again a week later. Such a test would have demonstrated reliability, or stability, over time.

A second type of reliability, the equivalent forms method of estimating reliability, involves administering two parallel forms of the test to the same group of people and correlating the results.

Another type of reliability, internal consistency, is considered an estimate of the degree to which a group of test items are performing their functions consistently. A test that shows a high level of internal consistency indicates that items on the test, or in a sub-test, are working together to measure the same thing, rather than many different things.

These considerations lead naturally to a question of validity, that is, what "thing" or "things" does a particular instrument actually measure? While reliability reflects the consistency of a measuring device, it is *validity that indicates whether a test measures what it is supposed to measure.* Though the presence of reliability makes it possible for a test to be valid, it does not guarantee the test's validity. For example, it is possible to have a bathroom scale that is consistently five pounds overweight; the scale is reliable--consistent--but is not a valid indicator of true weight.

In summary, when an instrument has demonstrated a high degree of reliability, an evaluator can be reasonably sure that it will measure consistently with different groups and at different times, which is especially important when an evaluation design calls for pre and post-testing or the use of comparison groups. When a test has proven validity, it can be trusted to measure accurately what it is intended to measure. The concept of validity depends upon the use to which a measure will be put. A test could, for instance, be a valid measure of mechanical reasoning, but if an evaluation is to assess mechanical ability, then the test is not appropriate.

As a final note, data on the reliability and validity of most published tests are available in the tests' manuals. Unless evaluators are fairly confident in their ability to interpret test manuals, which can be baffling indeed, it is advisable to seek the aid of a measurement specialist or test review books. Determining the extent to which tests developed by local school personnel are valid and reliable is slightly more complicated. Few local sites have the resources to conduct the required time consuming and expensive validity and reliability studies. Nonetheless, evaluators should be prepared to justify the use of locally-developed instruments, perhaps asking program personnel to review the instruments. Again, when in doubt, the assistance of a person trained in educational measurement would be of value.

Make the data collection as unobtrusive as possible.

All measurement alternatives should be implemented in a way that does not excessively intrude into the program's normal operation. Ideally, measurement devices are useful for instructional as well as evaluation purposes. The ultimate purpose of evaluation is to improve the quality of education for students, parents, teachers, and others, but this purpose can be defeated when the evaluation effort inordinately interferes with the learning/teaching process.

Instruments should be sensitive to positive as well as negative outcomes

In addition to picking up the positive outcomes that a program is designed to effect, measures should also be constructed to pick up information on possible detrimental program effects.

SECTION G

MEASUREMENT DEVICES

SECTION G

MEASUREMENT DEVICES

Dr. Nakamura's next task is to familiarize herself with the various measurement instruments which are available to answer the evaluation questions she has identified. Table 1 lists and describes the most commonly used measures: tests, reports, observations, and records. This section of the workshop discusses reports, observations and records. Section H focuses on the selection and construction of tests.

Section G briefly discusses concerns in the development of reports and observations schedules. For a more elaborate and comprehensive treatment of how to construct the various measurement devices discussed in this section, consult any of the references at the end of the section.

Table 1 - Measurement Devices

I. Tests

   A. Selected response tests (e.g., true-false, matching, multiple choice)

   B. Constructed response tests (e.g., essays, short answer, completion items)

   C. Work samples (e.g., examinee produces a product - paints a picture, sews a shirt, completes a job application form, writes a business letter, etc.)

   D. Process tests (e.g., a performance is judged as it occurs - reading aloud, playing piano, throwing a softball)

II. Reports

   A. Self-Report Devices

      1. Questionnaires and surveys

      2. Interviews

      3. Logs and journals

   B. Formal reports (e.g., annual reports submitted by a principal to the district office, status reports prepared by a program director)

III. Observations

   A. Structured (e.g., a checklist of specific behaviors which an observer uses to record student actions)

   B. Unstructured (e.g., a form on which a teacher periodically records impressions of the way classroom activities occur)

   C. Inventory (e.g., a checklist used by the school librarian to inspect program materials)

IV. Records (e.g., regularly maintained school records such as attendance records, minutes from meetings, cafeteria sales; classroom records indicating use of equipment, student progress, etc.)

REPORTS

Questionnaires

Questionnaires are an economical way to gather information from a wide audience since they can be mass produced and mailed to many people. If they are carefully constructed to provide useful information, they are valuable measurement devices. It is not easy, however, to develop a good and useful questionnaire. The content of a questionnaire is important because it must gather information the decision makers need. Its format is also important because it should motivate people to complete the questionnaire. A good strategy is to develop a clear, short questionnaire. It can include "open-ended" questions which ask the respondent to write in answers and selected response or "closed" questions which present a series of choices from which the individual checks off or circles a choice. The following example illustrates a combination of the two methods:

1. "Have you noticed the amount of violence on T.V.?"

Yes 1
No 2
Don't know 3

(If you answered 2 or 3, STOP here)

If "yes"

2. "How do you feel about the amount of violence on T.V.?"

Approve 1
Disapprove 2
Undecided 3

3. "Why do you feel this way?" ______________________________

____

____

(Example suggested by A. N. Oppenheim's Questionnaire Design and Attitude Measurement, Chapter 3).

While open-ended questions are easier to develop, they elicit a wide variety of answers, consequently, the information collected from these questions is harder to analyze. In contrast, closed questions are more difficult to construct, but are easier to analyze.

Before constructing the questionnaire, it is wise to conduct open-ended interviews with some program participants to determine some of the objectives that need to be measured or explored. A rough draft of the questionnaire should then be developed based on these interviews. This draft is given to a small group of program participants to see if the questions are finding out what the evaluator wanted to know and if the questions are clear enough that respondents are able to answer similar questions consistently. For example, a questionnaire may contain two questions asking whether students felt that a segment of instruction was clearly presented. If the answers to these questions are not similar, there may be a problem with the way the questions are phrased.

Interviews

Interviews are an alternative to questionnaires and are especially useful when the population to be questioned is fairly small. They are more expensive to administer than questionnaires because of costs associated with hiring and training interviewers. However, when administered by a trained interviewer they have certain advantages over questionnaires. The interviewer is able to go beyond the list of questions and collect other information that would not show up on a questionnaire. The interviewer can also probe responses by asking followup questions. However, it is important that the interviewer cover the same list of questions with each person interviewed. This consistency will ensure valid results and allow a comprehensive summary of the interviews.

OBSERVATIONS

Observations in evaluation can be informal or formal. Informal observations involve a casual look at a program by a few observers. They are quick and easy to carry out and are especially useful in small scale evaluation. Formal observations are more strictly controlled. They specify times of observations and methods for recording information.

Observations can take place in either "controlled" situations where people know they are being observed or in "natural" settings where people are observed in everyday situations and measurement is unobtrusive. In natural settings the evaluator does not manipulate conditions. There is a good possibility that the evaluator's presence is unnoticed, or if noticed, should not affect his findings since he is not having any direct interaction with those being observed. The evaluator can examine samples of behavior and pieces of conversation in such varied settings as assemblies, faculty and student lunchrooms, playgrounds, faculty meetings, PTA meetings, and recesses. For example, attending faculty and PTA meetings could help identify the kinds of praises, doubts, and constructive recommendations that exist about the program that is being evaluated.

RECORDS

Carefully kept school and class records can be one of the most substantive yet inexpensive sources of information. However, even when the records appear accurate and complete, an evaluator should try to utilize other measurement devices to substantiate the findings. For example, an evaluator ran into this situation midway through a Summative Evaluation of a Counseling Program. She discovered that many eleventh graders involved in the program were being reported absent from their study periods. Further investigation from interviews with students and teachers revealed that these same students had independently organized specific subject study groups during that period of the day. This gave a very different picture than the absenteeism records.

SECTION H

CHOOSING TESTS TO MEASURE ACHIEVEMENT OUTCOMES

SECTION H

CHOOSING TESTS TO MEASURE ACHIEVEMENT OUTCOMES

H. Choosing Tests to Measure Achievement Outcomes
  1. Considerations in Making Test Specifications
  2. Sample Test Specifications
  3. Exercise 4

As the previous section pointed out, there are a variety of measures Shelley Nakamura can use to answer her evaluation questions. Since tests are frequently used to measure outcome questions (especially dealing with cognitive and psychomotor objectives) this section focuses on their selection and construction.

Evaluators who are committed to using tests that directly and fairly measure achievement outcomes are faced with "over-choice." There are published tests, measures developed by program staff, tests that require examinees to develop their own answers, those in which examinees choose from among given alternative responses, and so on. On what basis should an evaluator's choice of a test be made?

No official tally has been made of the number of program staff members who have exclaimed upon reviewing disappointing test results -- "But the test didn't assess what we taught!" Nor has any record been kept of the decision-makers who have asked, "What does this really mean?" when presented with evaluation findings such as, "70% of the students achieved a score of 75% on the XYZ Test." Such concerns, though, are real, and they have highlighted the need for tests that are sensitive to instruction and which provide meaningful information about achievement of cognitive and psychomotor outcomes.

One approach to selecting or developing appropriate achievement tests is to outline, for each learning outcome to be measured, a set of test specifications. A test specification is a clear definition of the examinee behaviors to be measured and the context in which they will be measured.

There are two basic ingredients in a set of test specifications:

1) a description of the stimulus, that is, what do examinees encounter before making a response?
2) a description of the kind of response required of examinees, i.e., what are they expected to do during the test?

Making test specifications may at first seem to be the long way around for accomplishing the apparently straightforward task of choosing valid and reliable tests but it is a useful procedure for several reasons:

-- The process of detailing the stimulus and response elements can help ensure that both program staff and the evaluator share the same interpretation of the outcome to be measured. The fairness of the evaluation clearly is maximized when there is a match between what was taught and what was tested.

-- On the practical side, test specifications give the evaluator guidelines for reviewing the suitability of existing instruments

-- When a test is to be developed by the evaluator or staff, the specifications can serve as a "blueprint" for generating test items.

-- Clearly written test specifications help to remove the mystery from testing. They provide a way to make the evaluation open to public inspection by allowing staff, parents and others to see exactly what was measured.

-- The specifications can be used as a vehicle for communicating with decision-makers who will interpret examinee achievement on the basis of test scores.

The contents and format for test specifications can vary from complex to simple. As a rule of thumb, more detail will be needed when the specifications are to be used as a basis for developing tests rather than just reviewing tests, since the specifications must be complete guidelines for the test item writers. Above all, test specifications would effectively communicate to staff, test writers, decision-makers and other groups who have a stake in the evaluation exactly what is to be measured and in what context. Some of the questions that might be considered when developing test specifications are outlined on the next page, followed by a test specification "sample."

CONSIDERATIONS IN MAKING TEST SPECIFICATIONS

I. STIMULUS ELEMENT OF THE TEST: WHAT DO EXAMINEES ENCOUNTER BEFORE MAKING A RESPONSE?

A. Test content

- What is the range of subject-matter to be covered in the test?
- What concepts, rules, principles should be incorporated into the test?
- What skills or abilities are to be measured?
- What are the limits of the content of the test?

B. Test format

- How should the test content be presented?
- What should the test items look like?
- What directions should be given the examinees?
- What does a sample item look like?

C. Materials

- What materials, equipment, or supplies are needed for the testing procedure?

II. RESPONSE ELEMENT OF THE TEST: WHAT ARE EXAMINEES EXPECTED TO DO?

A. Form of response

- Should examinees choose an answer from among options given on the test ("selected response" items): multiple choice, true-false, matching, etc.?
- Should examinees construct their own responses?
  - Should they respond orally or in writing to questions? problems? statements? (e.g., essay tests, short answer completion items)
  - Should they provide a "sample" of work? (e.g., sew a blouse, fill out a job application form, make a wood construction)
  - Should examinee's performance be judged as it occurs? (e.g., dancing, reading aloud, throwing a softball)

B. Criteria of Correctness

- What constitutes a correct response?
- What constitutes an incorrect response?
- For selected response items, what are the "wrong answer" options?

III. ADMINISTRATION OF TEST

- How much time is permitted?
- Do administrators need special training?

Sample Test Specification

Evaluation Question: Can students distinguish between factual and slanted reporting?

I. Stimulus Element

A. Content

The learner will be given short news stories or factual accounts of events. The reports will concern national, world, local or school news events. Each account or story will be two to four sentences and will exhibit either objective reporting or slanted reporting.

Examples of slanted reporting will employ any one of the following reporting techniques:

- slanted language - using words or phrases which express approval or disapproval
- exaggeration - making things seem far greater or far less than they really are
- selective reporting - using information that only supports one side of the issue. Can best be detected by comparing two or more reports of the same incident or issue, and by paying attention to what the report doesn't say
- use of unreliable source - (person from whom a reporter gets information) a person who cannot be trusted to give accurate information and one who is not in a position to know important facts about an event
- misquoting - misrepresenting a person's statement(s) by changing what was said, and selectively deleting portions of the statement

Examples of objective reporting will not employ any of the slanted reporting techniques.

B. Test Format:

Directions: "In each of the following reports, decide whether the report is an example of factual reporting, or slanted reporting. Circle your answer."

Sample Item:

One of the most important events ever in the world of sports occurred today at Lansing High. Sheila Jones was the first girl allowed to play on the varsity basketball team.

a. factual reporting

b. slanted reporting

C. Materials

Individual test booklets for each student.

II. Response Element

A. Form of response

The learner will indicate whether a given news story or account is an example of objective or slanted reporting by circling the appropriate response.

B. Criteria of correctness

Items which examples of slanted reporting will exhibit one of the slanted reporting techniques (see I-A).

Items which the student indicates are examples of objective reporting will be free of any of the slanted reporting techniques.

III. Administration of test

The test should be taken in one class period.

Sample Test Specification

Evaluation Question: Can students identify concepts of time management when provided with a series of situations which apply these concepts?

I. Stimulus Element

A. Content

The learner will be presented with a list of five time management concepts and a list of 20 situations in which these principles are applied (four situations per concept).

Concepts are:

1. the concept of priorities
2. the concept of task analysis
3. the concept of behavioral consequences
4. the concept of planning
5. the concept of trade-off

Situations will be one sentence descriptions of an individual applying one of the concepts to manage his/her daily life.

B. Test Format

Directions: Read each of the following situations. From the list of time management concepts at the top of the page, choose one concept which applies to the situation. Write the number of the correct concept in the space next to the situation.

Sample situations:

1. George rank-ordered the things he had to do on Tuesday, setting the most important things first (applies concept of priorities).
2. Joe made a list of the things he had to do (applies concept of planning).

C. Materials

Individual test sheets.

II. Response Element

A. Form of response

The learner will write a single number in the space provided next to each situation.

B. Criteria of correctness

Learner will write number of appropriate concept for each example.

III. Administration

30 minutes will be needed to complete the test.

EXERCISE 4

CONSTRUCTING TEST SPECIFICATIONS

Directions: This exercise involves developing a set of test specifications to answer the following Evaluation Questions for the Consumer Education Program:

"Can students identify comparison shopping techniques?"

"Considerations in Making Test Specifications" on page H-4 and the sample specifications which follow it may be helpful in completing this exercise. The content limits of the test specifications and the response format have been set and are filled in on the exercise sheet. Work as a team to complete the rest of the specification using the information that is given.

Please bear in mind that the test specification format used for this exercise is representative, but not inclusive, of the way test specifications can be developed.

EVALUATION QUESTION: Can students identify comparison shopping techniques?

I. Stimulus Elements

A. Test Content

Learner will be given a sample of a specific product for consideration. This product will be familiar to the learner population, as well as being one of daily usage (e.g., soap).

The test item will consist of two statements - one true and one false. The true statement will apply one of the principles of comparison shopping. These principles are:

Pricing Items -- comparing relative prices of items

Quality Inspection -- (i.e., inspecting clothes for quality of construction)

Quantity Purchasing -- (i.e., buying three or four of the same item for a discount on total price)

Where to shop -- (i.e., discount versus specialty store)

The false statement will contain consumer education information (e.g., steps to follow in consumer complaints) other than comparison shopping principles. Students are to indicate which statement applies a comparison shopping principles.

B. Test Format

Directions:

Sample Item:

C. Materials

II. Response Elements

A. Form of response

Learner places a check mark in the space provided to the left of each statement.

B. Criteria of correctness

III. Administration

SECTION I

MEASUREMENT CONCERNS FOR EVALUATION QUESTIONS

SECTION I

MEASUREMENT CONCERNS FOR EVALUATION QUESTIONS

I. Measurement Concerns for Evaluation Questions

  1. Evaluation Questions
  2. Planning an Efficient Measurement Strategy
  3. Exercise 5

Now that Shelley Nakamura is familiar with the various measurement devices which are available, she must select a device for each evaluation question. Chart below suggests those devices that are usually most appropriate for measuring each type of question.

Measurement Devices Commonly Used for Evaluation Questions

Evaluation Questions	Suggested Measurement Instrument
1. Outcomes	
A. Cognitive and Psychomotor outcomes	Achievement and performance tests
B. Affective Outcomes	Questionnaires, Interviews, Observations
2. Implementation	Records, Observations, Questionnaires, Interviews, Reports
3. Cost	Charts, Questionnaires
4. Research	Tests, Questionnaires
5. Unanticipated Outcomes	Questionnaires, Interviews

Although these are often used measures for each evaluation question, there are particular characteristics for each type of evaluation question that will influence the type of measure that would be most appropriate to use. The remainder of this section will outline these characteristics.

EVALUATION QUESTIONS

Cognitive and Psychomotor Outcomes

Tests are generally used to measure cognitive and psychomotor outcomes. As discussed in Section H, test specifications facilitate selecting or developing tests to measure these outcomes. Too often evaluators focus on selected or constructed response tests to measure cognitive outcomes. When appropriate, work samples should also be used to test cognitive outcomes. Frequently work samples can be used to measure performance on a number of objectives. For example, if a student sews a shirt, ability to follow pattern directions, operate a sewing machine, tailor a pattern, can be measured by examining the final product. In the case of psychomotor objectives, process tests in which a performance is judged as it occurs are appropriate.

Affective Outcomes

Measuring affect -- attitudes, opinions, feelings, beliefs -- is not an easy task. Unlike a cognitive or psychomotor outcome (running the fifty yard dash, multiplying two digit numbers), a person's affective orientation is a state of mind. In other words, it is impossible to directly tap a person's feelings, opinions, or beliefs for measurement purposes. Instead, an evaluator can only infer that people hold certain attitudes through their words and actions, and only under particular conditions and situations. With these concerns in mind, what can an evaluator do to measure important affective outcomes?

In all instances, the evaluator, staff, and other interested groups need to determine what they will accept as "proof" of the attitude. Consider an affective outcome question: "Did students' attitude toward school become more positive?" Is it enough for students to say in a questionnaire or interview that they "like going to school?" Or would it be more convincing to determine if they voluntarily attended an after-school learning lab?

In most instances that involve measurement of attitudes, the evaluator's wisest course of action is to use at least two different measurement strategies. If results from multiple measures aimed at assessing the same affective state of mind are consistent, then the evaluator can have much more confidence in the results. If results are contradictory, then perhaps one -- or all -- of the measures are not adequately tapping the attitude, and still another strategy may need to be employed.

In addition, the use of at least one observational measure is highly recommended. To develop an observation instrument for measuring affect, the evaluator must ask "What behaviors or actions would be likely indicators of the attitude we want to measure?" The starting point for answering this question is to find a term which summarizes the desired attitude, such as "tolerance," "cooperation," "appreciation" and so on. Then, list the observable activities, performances, or behaviors that a person who holds the attitude would be expected to exhibit. Once refined, such a list comprises an observational sheet which can be used as a guide for inferring the extent to which the desired attitudes are present. For example, in an evaluation of the counseling program at a senior high school, one of the evaluation questions had to do with "positiveness toward school" arising from the counseling program experience. The measures taken included examination of such school records as attendance, unexcused absences from class, tardy slips and discipline slips from teachers. These measures were taken before and after the counseling program in the belief that an improved attitude toward school would be reflected in less absences, tardies and disciplinary reports.

Implementation Questions

Measurement devices for program implementation should help answer the question, "What does the program look like in operation?" Existing records, such as lesson plans, minutes from meetings, administrative documents, can paint a picture of program operation. It also is quite possible that classroom teachers or other program staff maintain records which can be used as a source of descriptive information on instructional processes. For instance, individualized instructional programs often rely on records to track student progress. Tabulating such data could indicate which program components were used, by whom, and with what results.

Observing and recording the events of key program sessions represents another implementation strategy. A thorough implementation effort could even include photography, tape recordings, and a collection of representative samples of student work.

Data not provided through records or observations can be obtained by using one or more report methods. A word of caution about using only questionnaires and interviews to document a program should be mentioned. Keep in mind that data collected through these two types of instruments are subject to the respondents' selective remembering (i.e., recalling only the "worst" or "best" part of a session), plus biases in favor or against particular aspects of a program operation. Accurate records and controlled observations are therefore generally more useful measurement strategies for recording implementation factors.

Cost Questions -- A Special Case

Measuring the "cost" of a program is both a challenging and controversial effort. For example, even sophisticated cost analyses performed by "blue ribbon" teams of economists, accountants, and management scientists often are challenged by groups having a stake in the program under evaluation. This state of affairs stems from the fact that there is no universally accepted definition or set of procedures for investigating cost, and the issue frequently becomes, in the final analysis, a matter of judgement. Ordinarily, it is the legitimate province of decision-makers (or those who commissioned the evaluation) to make judgements about the way in which data the evaluator collected ought to be interpreted, as different analytic schemes such as cost benefit analysis, benefits foregone, etc., result in different conclusions. The evaluator's main responsibility, unless otherwise charged, is to provide objective and credible data to those people who will, in turn, make educational decisions. Therefore, evaluators should be prepared to carry out limited but credible cost investigations that reflect the stages of the entire "life cycle" of a program. These stages include:

- planning stage -- refining proposed program goals, selecting materials, instructional strategies, etc.
- start-up stage -- what is done to install the program for the first time
- operational stage -- while the program is running as expected after the "bugs" have been worked out

What follows is one approach to measuring cost that evaluators may wish to apply to small scale evaluations. It is, though, by no means the only available approach.

Just as the outcomes of instruction must be operationally defined before they can be measured, so must the concept of "cost" be precisely defined. Usually this means breaking down the general term "cost" into more specific categories which, in turn, represent the units for cost analyses. One sample set of categories might include the following:

- personnel — teachers, administrators, aides, etc.
- space — classrooms, offices, etc.
- equipment and supplies — typewriters, buses, computers, paper, etc.

It is quite possible that decision-makers may not be concerned with all three categories; perhaps they are interested only in the use of space in a newly constructed modular classroom or the rate at which supplies must be replenished. Evaluators should establish the "target" categories for cost analysis as early as possible in the course of the evaluation. For example, data could be collected using the following cost indicators:

A. Dollar expenditures including salaries, price paid for equipment, heating, installation of facilities, etc. Excluded from dollar expenditures should be "fixed costs" which exist regardless of the program. Fixed costs might be, for instance, administrative salaries, normal building maintenence, textbooks that are used in both new and old programs, etc.

B. Costs measureable in dollars that are not actual expenditures. This includes "hidden" costs like depreciation of equipment, fringe benefits for staff members hired exclusively for the program, estimates on the cost of space used solely for program instruction, etc.

C. Time spent on the program as well as time normally devoted elsewhere but "given up" because of the program. This included the time spent by teachers, students, administrators, secretaries and others associated with program activities.

D. Costs that cannot be quantified, such as staff morale, community attitude, student satisfaction.

A relatively straightforward strategy for selecting measurement devices to collect cost data can involve displaying the "targets" of cost investigation (categories) against different indicators of cost, as illustrated, in the chart below. Only some of the squares have been filled in to serve as examples of measurement devices that might be employed.

Chart 2

Cost Indicators / Categories	A Actual dollar expenditure	B Calculated dollar expenditure	C Time	D Non-quantitative costs
1. Personnel			A "log" of time spent on planning lessons	Interviews with aides to assess morale
2. Space		Records of utility bills, school property taxes, maintenance costs		
3. Equipment/ supplies	Invoices of payment for materials, checklist of wear and tear and upkeep			Questionnaire to assess teachers' and students' satisfaction with materials and equipment

4. *Research Questions*

Since research questions usually include a planned variation in the program processes, these questions most often utilize tests as the measurement device. Therefore, the concerns are similar to those for measuring cognitive and psychomotor outcomes which were discussed in Section H.

5. *Unanticipated Outcomes*

In addition to the expected program outcomes, often the program may have unanticipated outcomes or effects. These unanticipated outcomes may be as important as planned outcomes. They are vital in validating a program so the evaluator needs to uncover and measure them.

One way to systematically measure unplanned effects of a program is to ask individuals, using questionnaires and interviews, if they have reasonable evidence of any other outcomes, either negative or positive, that they would attribute to the program. People who may reveal important information include both the people directly in the program, e.g., teacher, students, as well as peripheral people, e.g., other teachers, parents.

PLANNING AN EFFICIENT MEASUREMENT STRATEGY

It is quite likely that many of the instruments selected to answer the evaluation questions can be adapted to address more than one evaluation concern. In fact, as an overall measurement strategy, it is helpful to select instruments that can be used to answer more than one evaluation question at a time. Fewer instruments will minimize the potential problem of overtesting. An efficiently planned measurement strategy will also make more economical use of available resources (both time and money).

For example, a single questionnaire can serve multiple purposes: it can help provide implementation information for a program by asking teachers to provide information about activities, materials, and schedules; it also can assist with identifying unplanned effects by including open-ended questions such as:

> "Do you feel the ________ program has affected students' learning in ways you did not expect? Please give specific examples."
>
> "Do you feel the ________ program has affected the school's normal operation in ways you did not anticipate? Please give specific examples."

The same questionnaire could also investigate concerns about cost by asking respondents about wear-and-tear on materials or equipment, the amount of time spent on the program which ordinarily is devoted to other duties, etc. In most instances, research questions can be addressed through the use of instruments addressing outcome and implementation questions. In all, the evaluator's aim should be to plan a data collection strategy that is comprehensive as well as efficient.

EXERCISE 5

CHOOSING MEASUREMENT DEVICES

In the next exercise you will go through the same type of procedure as Shelley Nakamura in selecting the appropriate measurement devices for three of the evaluation questions of the Consumer Education Program.

Directions: Exercise 5 has two parts. In the first part, each Team Member will work individually to identify three different measurement alternatives that could be used to answer a single Consumer Education evaluation question.

In the second part of the exercise, Team Members work as a group to arrive at an efficient data collection strategy in which more than one evaluation question is addressed in a single instrument.

PART 1: (Individual work)

Each of the evaluation questions are identified by team member (A, B, or C). For your evaluation question, list three alternative measurement strategies that could be used to answer it. You might find it helpful to refer to the Measurement Devices chart on page I-7 and to skim through the text in this section that pertains to the category of evaluation question on which you are working. List the instrument and the intended respondent group.

PART II: (Team exercise)

The purpose of this section is to develop an efficient data collection strategy. This process involves refining the initial selection of data collection devices by reducing the number of different measurement instruments needed in order to obtain the same amount of worthwhile information.

Exercise 5

SECTION I

Choosing Measurement Devices

PART I

DESCRIPTION OF MEASUREMENT ALTERNATIVES

TEAM MEMBER "A":

Affective Question: "Do students value planning ahead as a way to handle their personal finances?"

1. ____________________
2. ____________________
3. ____________________

TEAM MEMBER "B":

Implementation Question: "Is there any variation in program components for different students or classes?"

1. ____________________
2. ____________________
3. ____________________

TEAM MEMBER "C":

Cost Question: "What is the cost of supplementary materials and activities, specifically the field trips and Consumer Beware Game?"

1. ____________________
2. ____________________
3. ____________________

* * *

PART II: FINAL SELECTION OF MEASUREMENT DEVICES

INSTRUMENTS	Evaluation Questions	Suggested Respondent Group
1.		
2.		
3.		
4.		
5.		

SECTION J

EVALUATION DESIGN

SECTION J

EVALUATION DESIGN

J. Evaluation Design
  1. Comparison Groups and Randomization
  2. The Most Useful Designs for Summative Evaluations
  3. Exercise 6

In the last section, you, as Dr. Nakamura, chose the measurement devices to be used in answering each of the evaluation questions for the Consumer Education program. Now, a design must be formulated for each measure. A design is a plan for collecting information about the effects of a program. This plan includes target audiences that will be involved as well as the time and manner in which the data from the measure will be collected. In other words, the "who" and "when" of the measurement device.

Choosing a proper design is important in lending credence to the evaluator's findings. A proper design will influence the kinds of statements that can be made about the effects of a program. For example, if results show significant differences between the pre and post measures, this difference can be attributed to the effects of the program.

Traditionally, design plans are formulated for tests, which is the measurement device most often used in answering achievement outcome questions and research questions. Designs for tests are usually more complex than those for other measures. Therefore, much of this section will discuss evaluation designs for tests. However, designs for other measurement devices will also be discussed later in the section.

Often in summative evaluation some comparative information is required. Therefore, a test designer should be aware of some basic considerations such as: Is a comparison group being used? and, Are the comparison groups randomized?

Comparison Groups and Randomization

Both these terms become important in the question of assignment of students to be included in the program. Many evaluation designs require assigning students to groups so that comparisons can be made between groups. In evaluation, these groups are typically called experimental groups and comparison or control groups. Students in the experimental group receive the full treatment of the program being evaluated. Students in the comparison group participate in an alternative program, often the regular or traditional program. To ensure that the student population of both groups is as equal as possible, a process of random selection is used.

In random selection, each student has an equal chance of being in any control or experimental group. This may be accomplished in several ways. For instance, one procedure entails the lottery system. After determining the number of students needed in both experimental and control groups, the names of the available students are placed in a hat and randomly drawn, assigning the first name drawn to one group, the second name to the second group, etc. Another method involves working with a master list of names in alphabetical order, selecting every other name on the list and assigning each one to the groups alternately.

The major advantage of random selection is that it provides a cross section of student ability, experience, and maturity. Since randomization assumes that any existing variables are distributed evenly throughout the groups, the success of a program cannot be attributed to any constant difference between the groups. Arguments to which non-randomized procedures may be open include: "The kids in Program X were smarter," or "Program X shows better results because many of the parents of the children in that group were scientists" or even "The control group started at a lower level than the experimental group and so it naturally ended up at a correspondingly lower level." All these hypothetical criticisms illustrate difficulties that could be avoided by implementing a random selection procedure.

Of course, given school constraints, an evaluator cannot always control these factors. For example, often intact classes have to be used for a program thereby eliminating the possibility of random assignment to groups. When a comparison group is made up of people who were not randomly assigned, it is called a non-equivalent comparison group.

Pretest

Another consideration for a test design should be: Is a pretest being given? A pretest gives the evaluator a good measure of student knowledge prior to the program. It ensures that the groups cannot already perform the terminal objective. If they can, the objective of the program obviously must be re-examined. If a comparison group is to be used, a pretest also serves the important function of checking that both groups are fairly equivalent and thereby comparable.

The Most Useful Designs for Summative Evaluation

As previously discussed, the basic factors to consider when deciding on a design for a test are:

(1) Are you going to compare the treatment group with another group receiving an already established program (comparison group)?

(2) Are your comparison groups randomized or non-equivalent?

(3) Is a pretest being given?

The following chart and discussion details four classical designs used in summative evaluation. The right hand columns of the chart list whether or not each design contains a pretest, comparison group, and randomization of the comparison group. Clearly, this chart does not give an exhaustive list of possible test evaluation designs. An evaluator may find that an evaluation situation does not fit the criteria for any of the designs listed, and will need to set up another design. Other designs can be found in almost any statistical text.

Name of Design	Control Group?	Random Assignment?	Pretest?
(1) Pretest-posttest control group	yes	yes	yes
(2) Posttest only control group	yes	yes	no
(3) Pretest-posttest non-equivalent control group	yes	no	yes
(4) Time series	no	no	yes

The four designs in the chart are described in detail below.

	Control Group?	Random Assignment?	Pretest?
(1) Pretest-posttest control group	yes	yes	yes

This is an excellent evaluation design because it permits a powerful test to be made between program X and the alternative.

Students are randomly assigned to form two groups; one group will get program X (experimental group "E") and the other will not. The group which does not get program X may get no program or may get an alternative program (control group "C"). It is important to check the pretest scores to make sure that the two groups were more or less equivalent.

At the end of the program, if the average (mean) posttest score for the C-group is significantly lower than that of the E-group, then the difference may be attributed to the effect of program X.

Example. Student writing skills had been declining over a period of years in a senior high school. The tenth grade department chairperson secured the cooperation of all the English teachers to try a new program emphasizing writing skills. During the summer, school counselors randomly assigned students to two groups. Both groups took a criterion-referenced pretest in September. Then, the control group proceeded to work with the traditional writing unit that stressed diagramming sentences. The experimental group used a new program that included keeping a diary in class. Near the end of the school year in May, both groups took a criterion-referenced posttest, and the difference in the results between the two groups was analyzed for significance.

	Control Group?	Random Assignment?	Pretest?
(2) Posttest only control group	yes	yes	no

In this design, students are randomly assigned to either program X or the control group. No pretest is given but both groups are measured at the end of the program, and these measures are compared for significant differences.

	Control Group?	Random Assignment?	Pretest?
(3) Pretest-posttest non-equivalent control group	yes	no	yes

This design is similar to Design (1) except that the groups were not formed by random assignment. It is a very useful design when the evaluator cannot randomly assign students to programs but must deal with intact classrooms. The pretest scores must be checked carefully to ascertain the similarity of the two groups. When students are selected for a program based on some especially high or low score on a measure, the pretest becomes especially important. For example, if you are using a remedial class for the treatment, you would want to use a similar remedial group receiving an alternative program for comparison. In these cases of extreme scores, the non-equivalent control group (C-group) may have to be obtained from another school or district since all students eligible in your own school are getting program X.

Example. Students who scored three or more years below grade level on a sixth grade mathematics pretest were assigned to a special remedial group at the beginning of the year. At another school in the same district, the teachers decided to keep students with low scores mixed in with students in the regular math program. Both groups of slow students had similar pretest scores. At the end of the sixth grade, both groups received a mathematics posttest to see which group of slow students had advanced the most.

	Control Group?	Random Assignment?	Pretest?
(4) Time series	no	no	yes

This design is most practical in the school setting when it may be impossible to randomly assign groups or find a comparison group. The students in the program are used as their own control group. The same measurement is made on the same group of students at regular interals at least three times before and three times after the program. By observing if program X seems to affect the trend in the results, it is possible to determine whether or not program X appears to have had an impact on the outcome measure.

Example. An elementary school attendance counselor wanted some evidence that her small group counseling techniques were effective with students who had a history of unexcused absences. By checking the existing records for each of the three months prior to the counseling sessions, she figured that the students were missing an average of six days per month. She compared this with the three months after the counseling treatment and found that the students were absent an average of only two days per month.

Counseling Program Test Design

Last year, in the counseling program, Shelley Nakamura outlined the test design strategies that she would use during the evaluation. As an example, she decided a test would be a good measure for one of the outcome evaluation questions in the counseling program. In formulating a design for the question, she wrote the following:

Note that Dr. Nakamura listed the question, the measure, the "who" and "when" of the measure, and then combined the information in an evaluation design. This same procedure can also be applied to measurement devices other than tests, such as reports, observations and records.

These measurement devices require a design different from those used for tests, though the evaluator is still concerned with the "who" and "when" of the measure. Unlike design plans for tests, other measurement devices do not necessarily require information about comparison groups, randomization, or pretests. Instead, the evaluator usually needs to decide only what groups of people (e.g., students, teachers, parents) should be given the measure to yield the most information about the evaluation question. For example, in the counseling program, Dr. Nakamura decided to use a number of different measurement devices. The following examples describe some design plans she formulated which made use of a variety of measures.

The outline used by Dr. Nakamura is only meant as a suggested format. Any plan that includes the important elements--evaluation questions, measures, "who" and "when" of the measures and design--is fine.

Counseling Outcome Question

Can students who have completed the counseling program identify concepts of time management better than students who did not take the program?

Measures

A. A matching test

Evaluation Design Questions

1) Who will be measured?	Students in the counseling program and a control group of students who have not yet taken the program
2) How will the groups be constituted?	Randomized
3) When will they be measured?	Before the beginning of the program and at the completion of the program

Recommended Evaluation Design

Compare the pre- and posttest performance of students assigned to take the course with the performance of the comparison group made up of students who are on the waiting list for next semester's scheduled program. Administer tests to control students in their homeroom on the same days as the tests are given to the experimental group in the counseling session.

Counseling Cost Question

What is the cost of the counseling program in terms of counselor time, for counseling and management functions?

Measures

A. Reports (Interviews)
B. Records (Time Sheets)

Evaluation Design Questions

1) Who will be interviewed?	Counselors
2) When (how often) will the interviews take place?	At the end of the semester
3) When will the records be examined?	At the end of the semester

Recommended Evaluation Design

Measure A. At the end of the semester, interview counselors to obtain information as to how much time counselors spent interviewing teachers, collecting class records, interviewing clerks, collecting student data from the school office, and arranging with Duplication Services for the production of necessary forms.

Measure B. At the end of the semester, examine counselor time sheets for total time spent by counselors in the group counseling sessions.

Counseling Implementation Question

Did the counselors implement the instruction of each component of the program exactly as prescribed in the program guides?

Measures

A. Observations
B. Records (Counselor Logs)

Evaluation Design Questions

1) Who will be observed?	Counselors conducting counseling sessions
2) When will the observation take place?	At the second and fourth session
3) Whose records will be checked?	The counselors
4) When will the records be checked?	At the conclusion of the last counseling session

Recommended Evaluation Design

Measure A. Two counseling sessions will be monitored by two observers (same observers in each session) equipped with an observation guide used to ascertain if the treatment for that session is being given as prescribed.

Measure B. After the last counseling session, counselor logs will be checked for any indications of deviation from the prescribed treatment.

EXERCISE 6

DETERMINING APPROPRIATE EVALUATION DESIGN

For this exercise, only three of the evaluation questions and measures have been pulled out of this year's evaluation of the Consumer Education Program. Working as a team, describe an appropriate evaluation design to use for each of the evaluation questions and measures listed below. (The evaluation design for the research question is provided as an example.)

Research Question

Evaluation Question #13

What is the effect of frequency of the Consumer Beware Game on the achievement of program objectives?

Measures

A. Same instruments used to measure outcome questions (no new measures required)

Evaluation Design Questions

1) Who will be measured? (What groups?)	Students enrolled in the Fall course. Students on the waiting list for Spring course.
2) How will the groups be constituted?	Randomized
3) When will measures be given?	Before the beginning of the program and at the completion of the program

Recommended Evaluation Design

Randomly assign two consumer education classes to the existing instructional plan in which the game is played four times per week for 15 minutes. Assign two other classes to a modified instructional plan in which the game will be played two times per week for 15 minutes. Finally, assign two classes to a plan in which the game is not played at all. Compare post tests of these groups for significant differences.

Outcome Question

Evaluation Question #1

Are students enrolled in the program better able to differentiate between the proper and improper steps to follow in resolving a consumer complaint than those not enrolled?

Measures

A. Tests (multiple choice and short answer tests)

Evaluation Design Questions

1) Who will be measured? (What groups?) ____________________

____

2) How will the groups be constituted?

nonequivalent ____________________

randomized ____________________

3) When will measures be given? ____________________

Recommended Evaluation Design

____

____

____

____

____

____

Implementation Question

Evaluation Question #8

What are the basic program components, and are they varied for any reason?

Measures

A. Observation (congruency of major categories of teaching strategies and materials)
B. Reports (interviews)
C. Records (course lesson plans)

Evaluation Design Questions

1) Who will be observed/ interviewed? ____________________
____

2) When (how often) will observations/interviews take place? ____________________
____

3) When will lesson plans be reviewed? ____________________

Recommended Evaluation Design

Measure A. ____________________
____

Measure B. ____________________
____

Measure C. ____________________
____

Cost Question

Evaluation Question #9

How does the cost of the supplementaty activities and materials, especially the cost of the field trips and the Consumer Beware game, compare with the cost of required and other elective social studies courses?

Measures

A. Reports (interviews)
B. Records (school accounting records)

Evaluation Design Questions

1) Who will be interviewed? ____________________

2) When (how often) will interviews take place? ____________________

3) When will records be examined? ____________________

Recommended Evaluation Design

____

____

____

____

____

SECTION K

DATA COLLECTION

Section K

DATA COLLECTION

Guidelines for Data Collection
  - Orientation
  - During Data Collection

Common Problems in Data Collection
  - Instruments
  - Design
  - Analysis
  - Reporting

Once the major components of the evaluation information system have been described and approved, the evaluator is ready to implement the evaluation plan. The first steps will involve making arrangements with the central district office and individual schools for any special assignment of students, or classes to particular programs or experimental treatments.

Prior to actual data collection the evaluator can institute procedures that will facilitate smooth data collection. This section will present guidelines for the data collection process, describe common problems which may arise, and suggest possible steps for avoiding or coping with the problems that do arise.

GUIDELINES FOR DATA COLLECTION

·Orientation. One of the most effective means of ensuring cooperation during the evaluation is to keep groups affected by data collection informed about, if not involved in, decisions regarding the planned measures and the schedule for their administration. If administrators, teachers, students and parents understand and agree with the purpose, procedures and logistics of testing, the quality of data collection is likely to be enhanced. The following outlines certain basic information that should be conveyed.

·*Inform groups of the purpose and requirements of the evaluation.* Often people feel threatened or uneasy about evaluations. These fears, frequently accompanied by resistance or hostility, can be partially allayed by clearly presenting the purpose of the evaluation, its rationale, and any mandated data requirements. It is often helpful to stress the notion that the program, its processes and materials, not individuals are being evaluated.

·*Inform or agree upon data collection schedule and procedure.* By discussing the data collection schedule and methods prior to their implementation, the evaluator may be able to avoid unforseen logistical problems. In addition, groups then know when to expect data collection and can plan their other classroom activities accordingly.

·*Establish timing and format of both formal and informal reporting.* Groups affected by the evaluation are frequently concerned about who will be receiving evaluation information. The evaluator can specify what information collected will remain confidential between the evaluator and program staff and what information will be documented for the funding agency in a formal report. Often the most utilized evaluation data is information transmitted in informal troubleshooting sessions between evaluators and program staff. Discussing the best format for presenting data at informal meetings prior to actual data collection can serve two purposes: (1) the evaluator will know what data analysis and displays will be comprehensible and useful; (2) the program staff will realize how the data collection can help them.

·*Make data collection as unobtrusive, economic and convenient as possible.* When program unit tests or tests prepared by the evaluator can be used by teachers for assessment and placement and by the evaluation for gauging program effectiveness, teachers may feel more positive about the utility of the testing. Also, by discussing the data collection schedule procedures with those people involved, the evaluator should be able to work out procedures which are convenient and reasonably unobtrusive.

·*Establish role and style.* Acceptance is apt to occur if the evaluator is perceived as a competent, objective facilitator of program effectiveness. Also, an open, honest, flexible style can elicit trust, cooperation and candor from respondent groups. It is particularly important for an evaluator to avoid an adversary image. The evaluator should not confuse and alienate audiences by using scomplex academic jargon, particularly jargon related to statistics.

·During data collection. Once data collection has begun the evaluation staff must anticipate and, hopefully, avoid problems.

·*Remind respondents about the schedule.* A short phone call or note reminding groups about a scheduled classroom visit or test administration can avoid the often encountered "Oh I forgot you were coming. We can't do that today."

·*Check that any required program processes have occurred.* Too often, scheduled testing is out of synchronization with program pacing. Clearly a scheduled unit test should not be administered by teachers or evaluation staff if the unit has not been completed. Similarily, the use of enrichment activities cannot be observed if those materials have not arrived.

- *Be unobtrusive.* For example, deliver on promises to disturb classroom functioning as little as possible by arriving for classroom observation before class begins, so activities are not disrupted. Arrange to do individual testing in a separate room or at the back of the room. Arrange discrete, efficient procedures for calling individuals or small groups of students from a classroom.

COMMON PROBLEMS IN DATA COLLECTION

A number of problems can, and frequently do, arise as the evaluation is conducted. Hopefully, following the above guidelines will avoid some of these problems. When problems do occur, parts of the data collection plan may have to be modified or even abandoned. On the following pages, problems relating to instruments, design, analysis, and reporting are discussed.

- Problems relating to instruments.
  - It is not uncommon to discover that test results are impugned due to *errors or oversights in test construction.* Learners may be confused by poor directions or bad items. Entire tests or sets of items may have been at the wrong difficulty level. Each of these problems could have been avoided by developmental testing of the instruments on a sample of learners - both masters and non-masters. A review of the instruments by teachers might also have avoided these problems.

-Test results may also vary if *the standard administration procedures are not followed*. This may be indicated by either wildly discrepant test results within presumably homogeneous groups or by observations or interviews. For example, a teacher may "warm up" students on practice items or give helpful hints on difficult items. If these hints have led to large fluctuations in test results between initially similar classes, the deviant data should be dropped. If there is no discernible effect on test scores, data may be kept. In either case the importance of standardization of procedures can be stressed in meetings with the problem groups.

-*Sometimes scoring guidelines are violated*. This can happen particularly with unit tests which are classroom administered and scored. Variation in applying scoring criteria could have been avoided, perhaps, by initial testing of the criteria with a sample of scorers or by a more systematic orientation of scorers. Different teachers may be stricter or more lenient with student paraphrasing or exact wording on fill-in questions. Occasionally students may have been allowed to score their own tests, resulting either in judgmental errors or questions of test credibility. In some cases tests may be re-scored by the evaluation staff, or, if necessary, dropped. Again, early detection of the problem can allow clarification of the importance and logistics of the scoring procedure so that future unit tests will be useable.

•Problems with design.

•Sampling. Sometimes, too, the sampling plan indicated by the evaluation design is not implemented. Often this occurrs due to logistical problems such as not enough students of each required ability level enrolled in the program. Sometimes the sampling plan is not implemented due to confusion about its importance, desirability or feasibility. This might be avoided by meeting with the personnel who are responsible for student assignments (prior to class scheduling.) If it's too late or if pretesting indicates that presumably comparable or representative groups are not comparable, then the situation can sometimes be salvaged by identifying comparable subgroups by means of particular analyses.

•Scheduling. The testing schedule can also be disrupted by logistical hangups such as program materials not arriving on time or different schools starting the program at different times. If these situations could not have been avoided by earlier tracking of materials or by discussing with the program staff the importance of coordinating starting dates, the evaluator must then re-examine the effects on the design. It may be that pretests can still be given at standard times, but other potentially disruptive school activities must be carefully documented. If all else fails, the design may have to be modified.

If tests must be given at different times due to uneven pacing or intervening school programs and holidays, careful documentation of program processes may determine if the test data is useable.

°Problems with analysis.

Many times intended audiences of evaluation reports complain that they cannot understand the data. The pre-program orientation meetings could have avoided this when data formats, preferably accompanied by examples, were discussed. Otherwise the evaluator can "tryout" samples of the analysis explanations and displays with sample members of the audiences. During evaluation planning the series of meetings of evaluation staff with program staff can provide a means for working out useable analytical schemes.

A second common problem in analysis is that the data are not timely. Hopefully an agreed upon schedule for reporting data has been discussed. Also, hopefully the evaluation staff meets that schedule or re-negotiates a mutually satisfactory schedule. Certainly modifications of testing schedules due to uneven pacing will require modification of data analysis. The evaluator must avoid the common pitfall of preparing such a complex analysis that it takes too much time resulting in a report that is not only late but is also difficult to understand.

°Problems with reporting.

Finally, a number of problems can occur with reports, formal or informal, that are prepared during the data collection period. Recurring complaints are: program staff do not attend reporting sessions; no one reads, understands or agrees with the reports; the reports are not on schedule. Here too, orientation sessions can provide information and motivation for program staff about the merits and payoffs of the evaluation information. Involving program staff in negotiating formats and content for the reports can avoid some of these problems. It may also be necessary to contact staff disenchanted with the reports or reporting session in an attempt to identify and solve problem areas.

SECTION L

DATA ANALYSIS

SECTION L

DATA ANALYSIS

L. Data Analysis
  1. Descriptive Statistics
  2. Tests of Differences

The last section brought Shelley Nakamura to the point of administering those measures chosen to answer the evaluation questions. In this section, we will proceed to that time when the program is completed and all the measuring devices have been administered. Now the job of the evaluator is to assemble all the data from the various measurement devices, and then present the information in a way that will provide meaningful answers to each evaluation question. The process begins with an examination of the procedures used in data collection and concludes with an examination of well the results provide the evaluation information called for.

Reviewing Data Collection Procedures

The evaluator's first task in the analysis is to review and report all data collection procedures that were carried out. Since summative evaluation *should* be conducted on an already established program, an assumption may be made that the program had been implemented as designated in the program guidelines. However, the evaluator should record any available information on program process deviations. This information may be valuable later to help explain program results. For example, Dr. Nakamura found low posttest scores on one objective in the counseling program in one classroom. Examination of the teacher's log showed the lessons on that particular objective were not given for various reasons thereby explaining the low posttest scores.

Although implementation information should be gathered, the main analysis concern of the summative evaluator is usually those measures which dealt with total program results (e.g., outcomes, cost).

Evaluation designs specify the manner in which data are to be collected, (the "who" and "when" of the measure) but many things happen that may cause deviation from the original plans. As a consequence, all deviations must be noted along with the extent to which such deviations qualify the data or limit their usefulness in some way.

Standard administration procedures of the measure are a must for the data to have been properly collected, both in terms of when the measures were administered and the method of administration. If, for example, a pretest is given to all students two weeks after instruction has begun, those data can still be useful (if the program is of long enough duration), because although it was administered late, it was given to everyone at the same time. If the pretest was given to some students on time and to some students late, the results would be less credible, because there would be no way of knowing what confounding factors had influenced performance on the measure. Other areas of scrutiny include the manner in which the instruments were administered and to whom. Results would not be valid, for example, if one group of students taking a posttest were doing so with a time limit while another group was allowed to complete the test without any time restriction. The data collector must also make note of those people who were excluded from a measure who shouldn't have been (e.g., students absent during the posttest) and those who were included and shouldn't have been (e.g., students who did not take the entire program but answered a questionnaire).

What Are The Appropriate Data Analysis Procedures?

Once the data is collected, what should the evaluator do with it? Three categories of statistical analysis will be discussed: Descriptive Statistics, Tests of Differences, and Tests of Relationships. This workshop will discuss only the more common and simple methods of data analysis. This discussion is only intended as an overview of the approach and is not meant as a complete course in analyzing data. More sophisticated and complete methods are given in statistics courses or may be read in the references presented at the end of the section.

DESCRIPTIVE STATISTICS

Of the three categories of statistical analysis to be discussed, descriptive statistics should first be compiled for all data. The first step is to "eyeball" the data. This enables the evaluator to ascertain if there are any peculiarities in *scores* (e.g., number correct on tests, responses from questionnaires). Any extreme deviations, either high or low, should be checked for signs that the measure was taken incorrectly, e.g., partially completed interviews, questionnaires, incorrectly scored tests. The extreme deviations that are a result of an error in administration should be removed before the body of data is analyzed further.

Now the evaluator has a mass of scores. These individual scores probably won't be of much value to anyone trying to read the evaluation report. Therefore, the next step is to reduce the raw data into some form that can be easily comprehended by decision makers.

The evaluator should probably first summarize the data from any tests used to measure the program's objectives. For example, one of the terminal objectives in the counseling program was: students will be able to identify concepts of time management. Therefore, all students were given a pre and posttest; Concepts in Time Management (CTM). Dr. Nakamura compiled the following summary of posttest scores. The results are displayed and discussed below:

CTM Posttest

Student	Score(X)	X — X	x^2
1	20	5	25
2	20	5	25
3	19	4	16
4	19	4	16
5	17	2	4
6	17	2	4
7	16	1	1
8	15	0	0
9	15	0	0
10	15	0	0
11	13	2	4
12	12	3	9
13	12	3	9
14	10	5	25
15	8	8	40
Total	228		187

Number of Students (N) = 15

Mean ($\overline{X}$) = 15.2

$$\overline{X} = \frac{EX}{N} = \frac{228}{15} = 15.2$$

Mode = 15

Median = 15

Standard Deviation (S) = 3.655

$$S = \sqrt{\frac{ECS-\overline{X}^2}{N-1}}$$

$$= \sqrt{\frac{187}{14}}$$

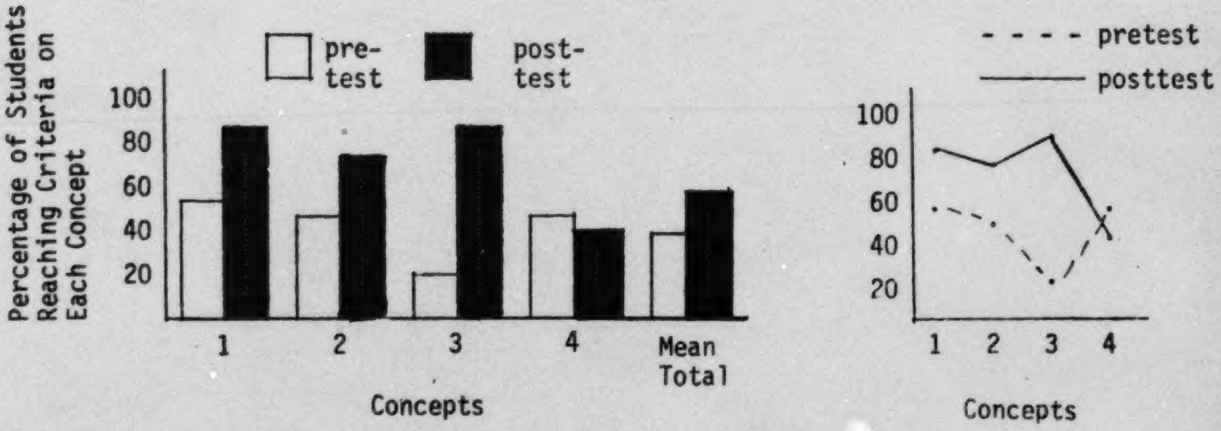
$$= \sqrt{13.357}$$

$$= 3.65$$

In the above analysis, Dr. Nakamura was answering the question often asked about a program's posttest, "What was the average score?" In statistical analysis, the average score can be either the mean, median, or mode. Each may give useful information and is easy to calculate. The mean is the statistically average score, the mode is the most frequent score, and the median is the score above and below which 50% of the students scored. In Dr. Nakamura's analysis of the Time Management test, the three m's were very close. This is not always the case. For example, sometimes there will be extreme high or low scores, which will cause the mean to be off center. In these cases, the mode or median may give a better picture of the overall performance.

Another statistic, the standard deviation, (S=3.655 in the above example) shows how much the scores are spread out around the mean or how much variability there is in the scores. The larger the standard deviation, the more spread out are the scores.

With the above information, the evaluator has a good idea how the scores "look." It is important to convey this information clearly to those reading the evaluation report. A widely recommended procedure for presenting data is to display the information in the form of tables and graphs. These provide an effective visual image of the data. In the analysis of the counseling program, Dr. Nakamura drew up a number of tables and graphs. In one graph, for example, she wanted to show the percentage of students who reached the criterion (4 out of 5 correct), for each concept on the CTM posttest as compared to the pretest. She compiled the following bar graph and frequency graph.



Percentage of Students Reaching Criterion on the Four Concepts of Time Mangement (Pretest and Posttest)

FIGURE 1

Notice how these graphs give much more meaning and life to the results than a prose description to the effect that:

> For Concept 1, 60% of the students achieved it on the pretest and 90% achieved it on the posttest.
>
> For Concept 2, 52% of the students achieved it on the pretest and 80% achieved it on the posttest.
>
> For Concept 3, 20% of the students achieved it on the pretest and 98% achieved it on the posttest.
>
> For Concept 4, 58% of the students achieved it on the pretest and 50% achieved it on the posttest.

Clearly, the graphs present a more easily understood summary of the data. The use of graphic presentation is eye-catching and may be the only part of an evaluation report that is reviewed by many of the individuals to whom the report is made available. Therefore, each graph or table must be clearly and fully labeled in order to be self explanatory. The use of a pictorial display, however, does not eliminate the need for a written description in the results section of the evaluation report. Each graph and table should be mentioned and described. It is important to remember that descriptive statistics give a clear picture of what happened in the measure, but do not show if there is any statistical significant difference or relationship between the scores. The importance of significance as far as enabling the evaluator to draw definite conclusions is discussed in the section dealing with Tests of Differences.

TEST OF DIFFERENCES

Once the scores of the group have been summarized on each measure, the summative evaluator will usually want to compare the performance of the group to that of another group that did not take the program. As discussed in the design section, the most desirable way to evaluate a program is to include a comparison group. Invariably there will be some difference in the performance of the two groups. For both groups to have identical scores on any measure is highly unlikely. Therefore, the evaluator can safely assume that the decision-makers will not be particularly interested in the fact that there was a difference but will want to know--"Was the difference significant?" That is, can the difference be attributed to the effects of the program rather than to chance. Anytime an evaluator can do a statistical test to find out the significance of the difference between two groups, it should be done.

Dr. Nakamura used a test of differences to compare the posttest performance of the students in the counseling program to the posttest performance of the comparison group. Since both groups had taken a pretest, Dr. Nakamura first ran a difference test to make sure that there was no significant difference between the two groups before the programs. The students had been randomly assigned so it was assumed there would be no significant difference, but since she had pretest scores anyway, Dr. Nakamura did a difference test just to be sure. Once satisfied that the groups were equivalent before the program, Dr. Nakamura then ran a difference test on the posttest scores of the two groups. The particular difference test that was appropriate for this design (pretest posttest randomized comparison group) is called a "t-test for unmatched groups." This is one of many similar difference tests. Each one makes certain assumptions about the number and kind of comparison or control groups.*

---

*For more information of the assumptions of various difference tests, consult any standard statistics book.

For the following test of differences, Dr. Nakamura used some of the descriptive statistics she had already compiled from the posttest scores of the experimental group (page 4) and comparison group (data not shown).

Experimental (E) Group	Comparison (C) Group
Number of Students (N_e) = 15	Number of Students (N_c) = 14
Mean ($\overline{X}_e$) = 15.2	Mean (X_c) = 11.0
Standard error (S_e) = 3.655	Standard error (S_c) = 2.99

$$t = \frac{\overline{X}-U}{S_{\overline{x}}}$$

$$t = \frac{\overline{X}_e - \overline{X}_c}{\sqrt{\left[\frac{(N_e - 1)S_e^2 + CN_c - 1)S_c^2}{N_e + N_c - 2}\right] - \left[\frac{1}{N_e} + \frac{1}{N_c}\right]}}$$

$$t = \frac{15 - 11}{\sqrt{\left[\frac{(15-1)3.65^2 + (14-1)2.99^2}{15+14-2}\right] \left[\frac{1}{15} + \frac{1}{14}\right]}}$$

$$t = \frac{4.2}{1.244}$$

$$t = 3.376$$

SECTION M

REPORTING INFORMATION

SECTION M

REPORTING INFORMATION

The results of the evaluation must be communicated to the appropriate audience. With a summative evaluation, this is generally done in the form of a written report at the conclusion of the evaluation. Occasionally, an evaluator is asked to provide an oral presentation as well.

The style and length of the report will vary from one situation to another, depending on the evaluation audience. Throughout the report the evaluator should use reporting strategies which are appropriate for the decision audience(s) who will be reading it. When there is more than one decision audience, as is frequently the case, the evaluator may prepare multiple reports, each report using reporting strategies suited to each audience. For example, an evaluator may be asked to make a formal report to a funding agency, a school board, teachers and parents. While a highly technical and detailed report is appropriate for a funding agency, a leaflet-like summary which explains the highlights of the evaluation in nontechnical terms is more suitable for parents. The report to the school board might include a nontechnical section which presents major evaluation results plus a more technical report presented as an appendix. By gearing the style and content of the evaluation report to the decision audience, the evaluator increases the likelihood that the report will be read and will influence people.

The following outline lists the points that should be covered in a report. Every point does not necessarily apply to every evaluation situation, but this list should help you avoid overlooking information that your audience will need in order to understand the results of the program and to have confidence in your conclusions and recommendations.

Outline of An Evaluation Report

	Suggested organization of the report	Suggested order in which the sections should be written
Section I.	Summary	last
Section II.	Purposes of the Evaluation	first
Section III.	Background Information Corcerning the Program	second
Section IV.	Description of the Evaluation Study	third
Section V.	Results	fourth
Section VI.	Discussion of the Program and Its Results	fifth
Section VII.	Conclusions and Recommendations	sixth

Section I. Summary

This section is a brief overview of the evaluation report. It summarizes the purpose of the evaluation and lists major conclusions and recommendations. Designed for the person too busy to read the full report, the summary should be no more than one or two pages long. Although the summary should appear at the beginning of the report, it is the section that is written last.

Section II. Purposes of the Evaluation

This section describes what the evaluation was intended to accomplish and what it was not intended to accomplish. In effect, this section is a description of the assignment that the evaluator accepted and should probably be prepared immediately after the evaluator accepts the assignment. A draft of this statement should be agreed upon by all interested parties and should be kept on file.

Section III. Background Information Concerning the Program

This section sets the program in context. It describes how the program was initiated and what it was supposed to do. If the evaluation audience consists of individuals who have no knowledge of the program, this section should be quite detailed. If, on the other hand, the evaluation report is to be read by people who are familiar with the program, Section III can be a brief setting down of facts "for the record."

A draft of Section III developed in the planning stages of the evaluation will insure that the evaluator has a clear grasp of the program, what it is and what it is supposed to accomplish. The draft could then be circulated to program personnel for their comments.

Typical content should include information on:

A. the origin of the program (the reasons for initiating the program)

B. the goals of the program

C. the characteristics of the program (materials and activities)

D. the faculty and others involved in the program

Section IV. Description of the Evaluation Study

This section describes the methodology of the evaluation. It includes the design for each evaluation question. If people are to have faith in the conclusions of the evaluation, they need to know in detail how the information was obtained.

A draft of this section should be written as the evaluation is being planned. It can then be circulated to the program personnel for their comments. Obtaining agreement beforehand as to what will constitute a fair measure of the program will increase the credibility of your results.

Section V. Results

This section presents the data that was collected from the various measurements. If the measurements were reliable and valid, these results are the "hard data." In addition, Section V may also include some "soft data"--anecdotal evidence or testimonials about the program.

The results section should not be written until all the data have been analyzed, recorded in tables, graphed or plotted, and tested for significance where appropriate. (Scores from tests are usually presented in tables showing means and standard deviations for each group. Results of questionnaires are frequently summarized on a copy of the questionnaire itself.)

Section VI. Discussion of the Program & Its Results

This section could be included in Section V where the results are presented. However, if the program and/or the evaluation is quite complicated, it is often preferable to have a separate section for interpreting and discussing the results. The results should be discussed with particular reference to Section II, Purposes of the Evaluation.

Typical content should include answers to the following questions:

A. How good were the results of the program? (How did the program results compare with what might have been expected had there been no program?)

B. How certain is it that the program caused the results? (Are there any other possible explanations of program results?)

C. What were the costs associated with the program--dollar and non-dollar costs (e.g., the using up of resources or the loss of alternative opportunities)?

D. What were the benefits associated with the program--dollar and non-dollar?

If possible, provide either at the beginning or end of this section a "Cost-Benefit Summary Table" in which costs and benefits, both dollar and non-dollar, are listed. Be sure to distinguish operating costs from start up costs, since the latter will not be needed if the program is repeated. And be sure to distinguish costs and benefits for which the evaluation has produced sound evidence from suspected costs or benefits which have not been substantiated by objective procedures.

Section VII. Conclusions and Recommendations

It is best to present this last section in the form of a list rather than in narrative form. Section VII is the most influential part of the evaluation report. Be sure, therefore, to emphasize what is important and to make clear what conclusions must be "tentatively" rather than "firmly" held. Take care that this section attends to all the concerns raised in Section II which described in detail the purpose of the evaluation.

Some evaluators feel that they should not make recommendations but should merely report evaluation data to decision makers. Others feel that the evaluator should make recommendations to decision-makers. In practice, the evaluator is frequently asked to make recommendations, those regarding the program and/or those regarding subsequent evaluations of the program (e.g., Should some instruments be modified or discarded for subsequent evaluations? Should a different design be used in future evaluation?) The evaluator has the ethical responsibility to make explicit the value base upon which recommendations are made. For example, an evaluator who believes in the merits of the developmental view of learning might make different recommendations than one who advocates objectives-based instruction. It is important that the evaluator clearly state the perspective from which recommendations are made.

The program planning team may also wish to make recommendations. Therefore, before writing this final section of the report, the evaluator might distribute copies of the results of the evaluation to the program planners. As a group they can discuss the implications of the results with the evaluator identifying those recommendations which are supported by the data. The planning committee's and evaluator's recommendations can then be incorporated in this final section.

MODEL REPORT

The sections you are about to read are only examples of the language and information included in a report. The information concerning results in the discussion and conclusion sections has been condensed so that only one evaluation question within each of the following major categories is discussed: program outcome, implementation, research questions and administration/cost.

I. SUMMARY (Will be handled in the next exercise)

II. PURPOSE

This report of the summative evaluation conducted for the Consumer Education program in Westlake Unified District will indicate whether or not the program should be continued in its present form, expanded or discontined. The report provides answers to the following specific evaluation questions:

A. Outcomes

(1) Are students able to list the steps to follow in resolving a consumer complaint?

(2) Are students able to list major local, state and federal consumer-related regulatory agencies and identify the functions of each?

(3) Can students demonstrate how to open and maintain a personal checking and savings account?

(4) Can students complete a state and federal income tax return?

(5) Can students list the advantages and disadvantages of buying on credit?

(6) Can students demonstrate a knowledge of comparison shopping techniques using price, quantity, and quality as variables?

B. Implementation

(1) Is there any variation in program components for different students or classes?

C. Cost Questions

(1) What can be done to keep down the cost of the field trips?

(2) What is the cost of running the program compared to other one-semester elective social studies courses?

D. Research Questions

(1) What is the effect of the "Consumer Beware" game on the achievement of course objectives?

E. Unanticipated Outcomes

(1) What avoidable and unavoidable problems have been caused by the field trips?

(2) Is the recent campus interest in student rights related to the Consumer Education class?

(3) What other outcomes do students and teachers attribute to the Consumer Education Program?

III. BACKGROUND INFORMATION:

A. Origin. The course was developed in response to requests from students, parents, and several board members that the high schools offer instruction in consumer affairs and accounting for personal use. As well as teaching the course, Bill Wolfson was the course developer. His developmental work over the course of one summer was supported with funds from the Office of the Assistant Superintendent for Instruction. It was agreed between the program funders and the program developer that a formal summative evaluation would take place during the program's third semester of operation.

B. Goals. The course objectives are stated in two broad categories, Consumer Awareness and Personal Accounting:

Consumer Awareness Goals

(1) the student will be able to list steps to follow in resolving a consumer complaint

(2) the student will list major local, state and federal consumer-related regulatory agencies and identify the functions of each

Personal Accounting Goals

(3) the student will demonstrate how to open and maintain a personal checking and savings account

(4) the student will complete a state and federal income tax return

(5) the student will list the advantages and disadvantages of buying on credit

(6) the student will demonstrate a knowledge of comparison shopping techniques using price, quantity, and quality as variables

(7) students will demonstrate a willingness to be systematic in handling their personal finances

C. Characteristics.

The course is one-semester (20 weeks) and meets five times per week during regular school periods of 50 minutes each.

SPECIAL MATERIALS INCLUDE:

(1) "Consumer Beware!" game adapted from a popular parlor game. The game is played several times a week and requires students to apply the skills they have been learning.

(2) Six consumer education filmstrips produced by the Food and Drug Administration, Federal Trade Commission, and Security and Exchange Commission. The filmstrips cover such topics as determining the quality of a product, calculating a standard unit cost for different package quantities, consumer protection laws and agencies, the function of banks, and different types of credit.

(3) Two speakers from the community. One speaker belongs to a consumer advocate group and explains what issues the group is currently pursuing and what courses of action

they are taking. The other speaker is a manager from the local credit bureau who talks about how individuals are assigned credit ratings and how credit can work for or against people.

(4) Three half-day field trips to a bank, car dealership, and supermarket.

STUDENT EVALUATION MATERIALS INCLUDE:

(1) Achievement tests of concepts

(2) Performance tests of skills (e.g., balancing a checkbook, completing a tax return)

(3) A term paper in which students select a current issue being handled by a regulatory agency, trace the history of the issue; describe the current status of the issue; and predict the resolution or nonresolution of the issue.

INSTRUCTIONAL MATERIALS:

A high school supplementary social studies textbook, American Consumer, and a variety of informational brochures from government agencies and private enterprise consitute the day-to-day course materials.

INSTRUCTIONAL STRATEGIES:

Lecture, class discussion, small group practice (Consumer Beware! game).

STUDENTS:

The course was designed for high school students (10th grade to 12th grade) of average ability. The course is an elective with priority for enrollment given to upper classmen. Students may not repeat the course unless they received a failing grade.

FACULTY:

The course is taught by Mr. Bill Wolfson (who also developed the course). Mr. Wolfson has a B. A. in economics and an M. A. T. from State University.

IV. DESCRIPTION OF THE EVALUATION STUDY

Outcomes:

To answer the evaluation question on achievement of outcomes, a multiple choice and short answer test was administered. The test was developed by the evaluator and reviewed by the teacher.

The evaluation design for assessing program outcomes used was a Control Group Pretest Posttest Design with random assignment. Pre- and post instruction performance of students in six classes who took the Consumer Education course was compared with the performance of students who did not take the course but were on the waiting list for the course in the Spring. The pretest was administered to the control students in their homeroom classes at the beginning of the year and the posttest was administered to them on the first day of Spring quarter classes.

Implementation:

An observation schedule listing the major categories of teaching strategies and materials was used to answer the implementation question. The observations were then compared with the lesson plan to reveal discrepancies. Follow-up interviews gathered information on the reasons for the deviations.

Cost:

Evaluation questions relating to quantative non-quantative costs were assessed by interviews with program administrators and by referencing school accounting records.

The principal and district business manager were interviewed about the relative costs of the Consumer Education program to the three required social studies programs and the three elective programs.

At the end of the Fall semester, school records were examined to compare quantative costs for the other six courses with those of the Consumer Ed program.

Research:

The research question on the effect of the Consumer Beware game on the achievement of program objectives was assessed by the same tests used to measure student achievement of program outcomes.

Students were randomly assigned to the six Consumer Ed classes. Two classes were then randomly assigned to each of the three treatment conditions involving use of the Consumer Beware game. Two classes did not use the game at all, two used it twice a week and two played the game four times a week.

Unanticipated Outcomes:

Questions relating to other outcomes were asked during the teacher interviews following the observations and in student interviews.

V. RESULTS

A. Program Outcomes. Achievement of program outcomes was measured by student performance on pretests and posttests. Outcomes 1, 2, 5 and 6 were assessed by multiple choice and fill-in tests. Work samples were examined for objectives 3 and 4. The pretest was administered the first day of instruction for the experimental group and the second day of instruction for the control. The control group took the test in their homeroom, supervised by the homeroom teacher.

Table 2 presents student performance on Outcome 1, "Are students able to list the steps in resolving a consumer complaint?"

		Pretest			Posttest		
	N	$\bar{X}$	SD	t-test	$\bar{X}$	SD	t-test
Experimental	124	22	5	NS	68	10	*
Control	37	21	6		19	5	

The two groups did not differ on the pretest but the experimental group had significantly higher posttest scores. Students in the Consumer Education program also exceeded the standard for mastery of the objective (80%) with their average score of 90%.

Similiar results occurred for the remaining program objectives.

B. Program Implementation. The question "Was there any variation in program components for different students on classes" was measured by observations of the classes at least four times each on two different days of the week. Each observation was followed by an interview with the teacher concerning any variations. Table 1 presents data gathered during the observations.

TABLE 1

Congruence of Observed Teacher Procedures with Lesson Plans

	1st Period	2nd Period	4th Period	5th Period	6th Period
1. percentaged activities done as scheduled	90%	87%	85%	92%	86%
2. number of activities not done as scheduled	31	32	36	28	33
3. activity omitted*	8	6	10	5	9
4. activities added**	7	9	8	9	8
5. activity sequence changed	16	17	18	14	16

*omitted activities were -- student reports (12 times)
-- discussion of homework (20 times)
-- filmstrip (6 times)

**activities added were -- speaker from mayor's office (5)
-- in-class exercise (27)
-- student reports (9)

Results indicate that for the most part program procedures were followed, few activities were omitted, changed in sequence or added.

Follow-up interviews revealed that omitted activities had usually been conducted on the previous day. The only activity that was really added, not shifted from another day's lesson plan, was a speaker from the mayor's office.

C. Cost Question. The school principal and the district business manager were interviewed to determine the costs for personnel, space, equipment/supplies of the consumer education class compared to other are semester required and elective social studies classes.* In addition, students in all social studies classes filled out satisfaction questionnaires the week before their final examinations.

---

*The required classes were American History, state requirements, and civics and environmental studies and European History were the elective classes.

TABLE 4

	A Actual Dollar Expenditure (Personnel)	B Actual Dollar Expenditure (Space)	C Actual Dollar Expenditure (Equipment/ supplies)	D Non-Quantitative Cost (student satisfaction)
A. Consumer Education	0	0	8%	0
B. Required Courses	15%	0	0	22%
C. Elective Courses	11%	0	3%	17.5%

The consumer education personnel costs were lowest for a number of reasons. This year was Bill Wolfson's first year as a salaried teacher which accounted for his receiving a lower salary than the teachers in the other programs; aides were used on the field trips, but they were parent volunteers (2 fathers and 3 mothers). The class required little extra equipment and supplies. The cost of equipment and supplies for the consumer education program was higher because of the field trips. Students' satisfaction was highest for the consumer education class.

D. Research Question. For the question on the effect of the "Consumer Beware" game on the achievement, two of the classes played the game twice a week, 2 classes played it four times a week and two classes didn't play the game at all.

TABLE 3

		Pretest			Posttest		
	N	$\overline{X}$	S.0	F-test	$\overline{X}$	S.0	F-test
4 times	65	23	6	NS	67	10	*
2 times	59	21	88		69	9	
2 times	55	22	6		50	2	

An analysis of variance F test revealed no significant differences among classes on the pretest and a significant effect on the posttest. Further analysis indicated a significant difference attributable to playing the game two times vs not at all, but no difference between playing it two times or four times a week.

Classroom observations indicated that the game was played similarly in all four classes.

E. Unanticipated Outcomes. The question "Is the recent campus interest in student rights related to the Consumer Education class?" was researched by documenting student's rights activities and by conducting interviews of 8 student rights leaders, who were identified by other students and staff. All documents related to student's rights (e.g., newspaper articles) were collected and a record of all activities (e.g., letter campaign to district state senator) was carefully kept. Student's rights activities/articles were reviewed and critiqued in terms of the similarity of these procedures with the consumer rights procedures taught in the course.

Results of the student interviews revealed that 6 of the 8 student rights leaders had taken the consumer education class and 5 students attributed their initial interest and expertise in student's rights to the class. They reported that they applied 5 specific consumer rights procedures learned in the class.

(1) mass appeal to public official (letter campaign to the state senator)
(2) communication with established pressure groups (wrote to the state university/student legislative lobbyist for information)
(3) communication with the media (articles were placed in both the Kennedy High and Maples High school newspapers)
(4) delineating existing rights (done in newspaper column)
(5) delineating legal recourse when rights are violated (done in newspaper column)

The student-generated list of applied procedures matched those consumer rights practices taught in the consumer education class.

EXERCISE 8

DRAWING CONCLUSIONS AND MAKING RECOMMENDATIONS

Directions: Part I

In this exercise, your team will refer to the results of the Consumer Education Program summative evaluation to draw conclusions and/or recommendations regarding three of the five categories of evaluation questions: program outcomes, program implementation, and the research question.

After each team member has read the following Results Section of the Model Evaluation Report for each of those evaluation questions, they are each to list one or more conclusions and recommendations the evaluator could make.

Team members can then discuss their individual suggestions to arrive at a team consensus.

Directions: Part II

Based on the Evaluation Report and the conclusions and recommendations reached in the first part of the exercise, each team member should write a Summary of the evaluation that will be appropriate for parents and businessmen in the community. The summary should be ½ to 1 page in length and should include a description of:

the program and its goals
the evaluation
the evaluation results and recommendations.

Exercise 8

Drawing Conclusions and Making Recommendations

PART I

OUTCOME QUESTION

Conclusions____

Recommendations____

IMPLEMENTATION QUESTION

Conclusion____

Recommendation____

RESEARCH QUESTION

Conclusion____

Recommendation____

Exercise 8

Drawing Conclusions and Making Recommendations

PART II

SUMMARY

SECTION N

FEEDBACK

FEEDBACK: SECTION A

EXERCISE 1: CATEGORIZING EVALUATION ACTIVITIES EXERCISE

1. Summative Evaluation
2. Needs Assessment
3. Summative Evaluation
4. Formative Evaluation
5. Program Planning
6. Formative Evaluation

FEEDBACK: SECTION C

EXERCISE 2: DETERMINING THE CONTEXT OF THE EVALUATION

Concerns About Program Characteristics

1. What is the most up-to-date source of information about the program's operation and goals?

   file on the Consumer Ed Program which contains the formative evaluation report, correspondence and an updated version of the program plan; Bill Wolfson

2. What activities do the staff feel are contributing most to program success?

   the Consumer Beware Game, practice exercises

3. Who are the individuals operating the program?

   the program teacher, Bill Wolfson
   the high school principal
   several parents

Concerns about Evaluation Requirements

4. What unexpected positive effects does the program seem to promote? What negative effects?

   unexpected positive effects - students' newspaper column
   - awareness of students' rights, especially the right to have open records and files

   unexpected negative effects - the Consumer Ed field trips are causing absences in other teachers' classes

5. What kinds of evaluation information are available or obtainable?

   observations, student questionnaires (given to students who were formerly in the program and who are presently in the program), and on the waiting list for the program

6. What monetary or staff resources are there to conduct the evaluation?

   Betty Washington will consult and her office will take care of clerical and data processing needs

Concerns about School Organizational Structure

7. Who are the decision makers?

   assistant superintendent
   district's high school principals

8. What are the lines of authority and communication within the program?

   direct to the assistant superintendent

9. Who are the potential critics of the program or of the evaluation? What are their concerns?

   two or three school board members who are concerned about the program's cost; other teachers who are inconvenienced by the scheduling of field trips

EXERCISE 3: FORMULATING EVALUATION QUESTIONS

August 26, 19xx

Mr. Philip McNeil
Assistant Superintendent for Instruction
Westlake Unified School District

Dear Phil,

As you know, I was extremely pleased that we were able to come up with a final set of evaluation questions yesterday at our meeting with Betty Washington, Irma Gruen, and Bill Wolfson. I feel it is a set of questions that will provide you with a sound data base for making a determination about the program. For the record, between now and the end of March, I will be collecting information on reporting answers to the following questions:

Outcomes

1) Are students able to list the steps to follow in resolving a consumer complaint?

2) Are students able to list major local, state and federal consumer-related regulatory agencies and identify the functions of each?

3) Can students demonstrate how to open and maintain a personal checking and savings account?

4) Can students complete a state and federal income tax return?

5) Can students list the advantages and disadvantages of buying on credit?

6) Can students demonstrate a knowledge of comparison shopping techniques using price, quantity, and quality as variables?

7) Do students show a greater willingness to be planful in handling their personal finances?

Implementation**

1) Is the Consumer Education game being used as planned?

2) Is there any variation in program components for different students or classes?

**I will be able to summarize background information on the origin of the program; individuals served by it and individuals operating the program by referring to Betty's files.

Cost Questions

1) What is the cost of running the program compared to other one-semester elective social studies courses?

2) What can be done to keep down the cost of the field trips?

Research Questions

1) What is the effect of field trips on achievement of course objectives?

2) What is the effect of the "Consumer Beware" game on the achievement of course objectives?

Unanticipated Outcomes

1) Is the recent campus interest in student rights related to the Consumer Education class?

2) What avoidable and unavoidable problems have been cuased by the field trips?

3) What other outcomes do students and teachers attribute to the Consumer Education Program?

Our next step is to develop the evaluation design and data collection instruments. I hope to be reporting to you on that by the time school opens.

Sincerely,

Shelley Nakamura

cc: Irma Gruen
Betty Washington
Bill Wolfson

FEEDBACK: SECTION H

EXERCISE 4: CONSTRUCTING TEST SPECIFICATIONS

EVALUATION QUESTION: Can students identify comparison shopping techniques?

I. Stimulus Elements

A. Test Content

Learner will be given a sample of a specific product for consideration. This product will be familiar to the learner population, as well as being one of daily usage (e.g., soap). Each test item will consist of two statements - one true and one false. The true statement will apply one of the principles of comparison shopping.

These principles are:

Pricing Items -- comparing relative prices of items

Quality Inspection -- (e.g., inspecting clothes for quality of construction)

Quantity Purchasing -- (e.g., buying three or four of the same item for a discount on total price)

Where to shop -- (e.g., discount versus specialty store)

The false statement will contain consumer education information (e.g., steps to follow in consumer complaints) other than comparison shopping principles. Students are to indicate which statement applies the comparison shopping principle.

B. Test Format

Directions: "Indicate, by placing a check mark, those statements concerning the purchase of product 'X' which apply the principles of comparison shopping."

Sample Item:

____1. If you wanted to find out about the quality of product X you would contact the Consumer Protection Agency for information.

_✓_2. Product X, found in both a specialty store and a discount store, would most likely be cheaper in the discount store.

_✓_3. It is wise to consult newspaper advertisements to price product X.

C. Materials

One test sheet per student

II. Response Elements

A. Form of response

Learner places a check mark in the space provided to the left of each statement.

B. Criteria of correctness

See 1-B

III. Administration

10 minutes is needed to complete the test.

FEEDBACK: SECTION I

EXERCISE 5: CHOOSING MEASUREMENT DEVICES

PART I

	DESCRIPTION OF MEASUREMENT ALTERNATIVES
TEAM MEMBER "A": Affective Question: "Do students value planning ahead as a way to handle their personal finances?"	1. questionnaire 2. interview 3. observation
TEAM MEMBER "B": Implementation Question: "Is there any variation in program components for different students or classes?"	1. observation 2. questionnaire/survey/interview 3. records
TEAM MEMBER "C": Cost Question: "What is the cost of supplementary materials and activities, specifically the field trips and Consumer Beware Game?"	1. records 2. formal report 3. interview

* * *

PART II: FINAL SELECTION OF MEASUREMENT DEVICES

INSTRUMENTS	Evaluation Questions	Suggested Respondent Group
1. questionnaire	affective	*student in the program/control group/teachers*
2. observation	implementation/affective	*Consumer Ed. classes*
3. records	implementation/cost	*teacher/school*
4. interview	cost	*school administrator*
5. interview	implementation	*Consumer Ed. teacher*

FEEDBACK: SECTION J

EXERCISE 6: DETERMINE APPROPRIATE EVALUATION DESIGN

The following designs are appropriate to use to answer the outcome, research, and implementation questions presented in the team exercise.

Outcome Question

Evaluation Design Questions

1) Who will be measured? (What groups?)	Students enrolled in the fall course. Students on waiting list for spring course
2) How will the groups be constituted?	
nonequivalent	
randomized	✓
3) When will measures be given?	pretest--fall; posttest--end of program

Appropriate Evaluation Design

Compare the pre- and post-instruction performance of students who have been randomly assigned to take the course in the fall with the performance of the students who are on the waiting list for the course in the spring semester. Administer the pretest to control students in their homerooms at the beginning of the year. Administer the posttest to control students on the first day of classes in the spring term.

Implementation Question

Evaluation Design Questions

1) Who will be observed/ interviewed?	observed--all classes interviewed--classroom teachers
2) When (how often) will observations/interviews take place?	observe--*four times, 2 days* interview--*end of each observation*
3) When will lesson plans be reviewed?	*each day of observation*

Appropriate Evaluation Design

Measure A. Observe all classes *four times each, on two different days of the week.*

Measure B. Interview classroom teachers *at the end of each day of observation* concerning the extent to which the basic lesson plan was followed and when and why it was varied.

Measure C. *On each day of observation,* review lesson plans for description of basic instructional sequences and activities.

Cost Question

Evaluation Design Questions

1) Who will be interviewed?	school principal, school administrator
2) When (how often) will interviews take place	end of fall semester
3) When will records be examined?	end of semester

Appropriate Evaluation Design

Measure A. *At the end of the fall semester, interview the school principal and the school administrator about supplementary activities and materials of the required social studies courses (e.g., World History, American History, Civics) and elective social studies courses (e.g., Consumer Education, Environmental Studies, Counseling).*

Measure B. *Review school records at the end of the semester for quantitative costs of the six courses, comparing categories of expense as it pertains to supplementary activities and materials for each course separately and for the required versus the elective courses.*

FEEDBACK: SECTION L

EXERCISE 7: DETERMINING APPROPRIATE DATA ANALYSES FOR EVALUATION QUESTIONS

1. b
2. a
3. a
4. b

FEEDBACK: SECTION M

EXERCISE 8: DRAWING CONCLUSIONS AND MAKING RECOMMENDATIONS

PART I.

OUTCOME QUESTIONS

Conclusions (THE PROGRAM PROMOTES STUDENT MASTERY OF ITS OBJECTIVES)

Recommendations (THE PROGRAM SHOULD BE CONTINUED, COULD BE USED IN OTHER SCHOOLS)

IMPLEMENTATION QUESTION

Conclusion (THE PROGRAM ACTIVITIES WERE GENERALLY CONDUCTED AS PLANNED)

Recommendation (SINCE PROGRAM OBJECTIVES ARE ACHIEVED WITH THESE PROCEDURES, THEY SHOULD BE ADHERED TO IN FUTURE USES OF THE PROGRAM)

RESEARCH QUESTION

Conclusion (THE CONSUMER BEWARE GAME IS EFFECTIVE ALTHOUGH USING IT TWO VS. FOUR TIMES PER WEEK MAKES NO DIFFERENCE)

Recommendation (USE THE GAME TWICE A WEEK)

PART II.

SUMMARY

The new program on Consumer Education initiated in Westlake District this year has been a success. The program was designed to teach students to be intelligent consumers, to know how to resolve consumer complaints and to manage their personal finances successfully. An evaluation of the program was conducted to determine if the program objectives were achieved, if its activities were used as prescribed, if a practice game was effective, and if the program was cost-effective.

The evaluation results indicate that the program was both effective and economical, so the evaluator has recommended that the program be continued. We look forward to the Consumer Education program in other district high schools soon!

SELECTED REFERENCES

Airasian, Peter W. Designing summative evaluation studies at the local level. In *Evaluation in education: Current applications*. Berkeley, California: McCutchan Publishing 1974

Anastaie, A. *Psychological testing*. New York, MacMillan Publishing Co., Inc. 1976

Babbie, E.R. *The practice of social research*. Belmont, California: Wadsworth Publishing Company, Inc., 1973

Buros,O.K. *Mental measurements yearbooks*. Highland Park, New Jersey: Gryphan Press

Campbell D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.) *Handbook of research on teaching*. Chicago: Rand - McNalley, 1963, pp. 171-246

Campbell D.T. Reforms as experiments. *American Psychologist*, 1969, 4, pp. 409-429

Center for the Study of Evaluation. *Elementary school evaluation kit: needs assessment*. Boston, Massachusetts: Allyn & Bacon, Inc. 470 Atlantic Avenue

Center for the Study of Evaluation. *CSE-RBS test evaluations: Tests of higher-order cognitive, affective, and interpersonal skills*. Los Angeles: Center for the Study of Evaluation, University of California, 1972

________. *CSE elementary school test evaluations*. Los Angeles: Center for the Study of Evaluation, University of California, 1970

________. *CSE-ECRC preschool/kindergarten test evaluations*. Los Angeles: Center for the Study of Evaluation, University of California, 1971

________. *CSE secondary school test evaluations*. Los Angeles: Center for the Study of Evaluation, University of California, 1973

Cochran, William C. *Sampling techniques*. New York. John Wiley, 1953

Cronbach, Lee J. *Essentials of psychological testing*. 3rd ed., New York: Harper & Row, 1970

Ebel, Robert L. *Essentials of educational measurement*. Englewood Cliffs, New Jersey: Prentice-Hall, 1972

Experimental designs for school research. Vimcet Associates, Inc., P.O. Box 24714, Los Angeles, California: Filmship-tape programs

Glass, Gene V. and Julian C. Stanley. *Statistical methods in education and psychology*. Englewood Cliffs, New Jersey: Prentice Hall, 1970

Haller, Emil J. Cost analyses for educational program evaluation. *In evaluation in education; current applications*. Berkeley, California: McCutchan Publishing, 1974

Hays, William L. *Statistics for the social sciences*, 2nd ed., New York: Hall, Rinehart and Wenston, 1973

Holland, James G., and Judith Doran. Instrumentation of research in teaching. In *Second handbood of research on teaching*, edited by Robert M.W. Travers. Chicago: Rand McNally, 1973

Lake, Dale G.; Mathew B. Miles; and Ralph B. Earle, Jr., eds. *Measuring human behavior*. New York: Teachers College Press, Columbia University, 1973

Metfessel, N.S., and W.B. Michael. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. *Educational and psychological Measurement 27* (1967): pp.931-943

McCall R.B. *Fundamental statistics for psychology*. New York: Harcart Brace Jovanovich,Inc., 1970

Oppenheim, A.N. *Questionnaire design and attitude measurement*. New York: Basic Books, Inc., 1966

Popham, W. James, *Educational criterion measures*. New York: American Book, 1971

Popham, W. James. *Educational evaluation*. New Jersey: Prentice-Hall, Inc., 1975

Popham, W. James, and Kenneth A.Sirdnik. *Educational statistics: Use and interpretation,* 2nd ed. New York: Harper & Row, 1973

Rosenshine, Barak, and Norma Furst. The use of direct observation to study teaching. In *Second handbood of research on teaching*, edited by Robert M.W. Travers. Chicago: Rand McNally , 1973

Siegal, Sidney. *Nonparametric statistics for the social sciences*. New York: McGraw-Hell, 1956

Simon, A., and E. G. Boyer, eds. *Mirrors for behavior: an anthology of classroom observation instruments*.. Philadelphia: Research for better schools, 1968

Sjogren, D.D. Measurement techniques in evaluation. *Review of Educational Research 40* (1970): pp. 301-320

Tuckman, B.W. *Conducting educational research*. New York: Harcourt Brace Jovanovich, Inc., 1972

Webb, E.J.,Campbell, D.T., Schwartz, R.D. and Sechrest, L. *Unobtrusive measures: non-reactive research in the social sciences*. Chicago: Rand McNally, 1966